UN GROUPE D'INDIVIDUS DÉCRIT PAR UNE VARIABLE QUANTITATIVE

DÉCRIRE LE GROUPE

SITUER LES INDIVIDUS DANS LE GROUPE

SITUER LE GROUPE PAR RAPPORT À UNE VALEUR DE RÉFÉRENCE

Mots-clés : Variables; Distribution de Laplace-Gauss ; Moyenne ; Médiane ; Mode ; Écart Absolu Moyen ; Écart-type ; Quantiles ; Quartiles ; Déciles ; Centiles ; Intervalle interquartile ; Score z ; Écart centré-réduit ; Rang percentile ; Échantillonnage au hasard ; Inférence ; Étapes descriptive et inférentielle ; Test *t* de Student de comparaison d'une moyenne à une valeur de référence ; Intervalle de confiance ; Inférence bayésienne.

Ce document a été établi en indiquant comment obtenir les différents résultats avec le logiciel SES-Pegase (version 7). Cependant, il a été rédigé dans le but d'être utilisé comme guide méthodologique et d'interprétation, quel que soit le logiciel utilisé.

TYPE DES DONNÉES ANALYSÉES

Nous présenterons l'analyse d'un dossier particulier, le dossier NOTEBAC. Mais cette analyse s'applique à toutes données de la forme suivante (Tableau 1) :

Tableau 1 : Structure du tableau des données individuelles

INDIV	X
i1	13
i2	10
i3	8
i4	11
i5	20
i6	20
i7	13
i8	7
(…)	(…)

Des informations ont été recueillies sur un ensemble (INDIV) d'individus (i1, i2…). Ces individus peuvent être, des personnes, des pays, des animaux, des voitures…
Parmi les données recueillies, on a une variable quantitative (X) observée sur les individus : une note, un temps, leur âge, leur taille, leur salaire, leur QI…

Questions

- Quelles sont les caractéristiques du groupe ?
- Quelle est la position du groupe par rapport à une valeur de référence ?
- Quelle est la position d'un des individus par rapport au groupe ?

UN EXEMPLE : LE DOSSIER NOTEBAC

Source : Denis Corroyer – Université Paris Descartes – denis.corroyer@parisdescartes.fr

On s'intéresse aux réponses à une question posée à des étudiants en 3$^{\text{ème}}$ année de psychologie à l'université Paris Descartes (Paris 5) : « Quelle a été votre note à l'écrit de français du baccalauréat ? »
La variable NOTE rapporte les notes des étudiants interrogés.
Les étudiants étaient invités à se choisir un pseudonyme pour pouvoir, s'ils le souhaitaient, se situer dans les tableaux et graphiques.
Le questionnaire a été rempli par tous les étudiants. Les données ci-dessous sont celles d'un échantillon de 34 questionnaires, extraits au hasard parmi cette population plus vaste.

On considérera qu'un écart est faible s'il est inférieur à 1 point et fort s'il est supérieur à 2 points.

Tableau 2 : Les données NOTEBAC

ETUDIANT	NOTE
SRP	9
EMA	11
SABX	8
EMIC	8
AMELIEM	8
CHEBLI	10
FUERTES	18
SCARLETT	11
CENIE	14
MATAHARI	11
CYCO	12
LUNE	12
LIDO	12
LOLA	11
ALBATOR	5
BIBI	10
MICKA	7
DR007	15
LOL	4
AGOAK	12
CLEM	7
MELB	11
NOUNOU	12
PAQUERETTE	11
SANDRAS	9
3D	11
AYA	11
SCOUBIDOU	11
ZAZ	8
JOE	14
MIKY	9
LOULOU	10
ZORRO	8
MELIS	10

Questions

Quelles sont les caractéristiques du groupe du point de vue des notes au baccalauréat ?
Quelle est la position d'un étudiant particulier (ZORRO) par rapport au groupe ?
Quelle est la position de ce groupe par rapport à la moyenne de référence, 10/20 ?

OUVRIR LE FICHIER

```
SES-Pegase:
Lancer SESAnalyse
Menu Fichier
– Ouvrir un dossier (*.SES)
Sélectionner le fichier NOTEBAC.SES
```

DÉCRIRE LE GROUPE

```
Menu Nouvelle analyse
- Sélectionner la variable NOTE en tant que "Variable(s) à analyser"
- Valider en cliquant sur OK
```

Une des premières choses à faire, avant tout calcul d'indices (moyenne…) est de représenter, sous forme graphique, l'ensemble des valeurs observées.

Analyser la forme de la distribution

Analyser la forme de la distribution consiste à s'interroger sur plusieurs points.
On se demandera si la distribution :
- est unimodale ou multimodale,
- est symétrique ou non,
- peut être ajustée par une distribution connue (distribution normale, distribution uniforme…).
On cherchera à détecter :
- d'éventuels sous-groupes de valeurs observées,
- d'éventuelles valeurs atypiques.

```
Menu Statistiques
– Distribution
- Histogramme
```

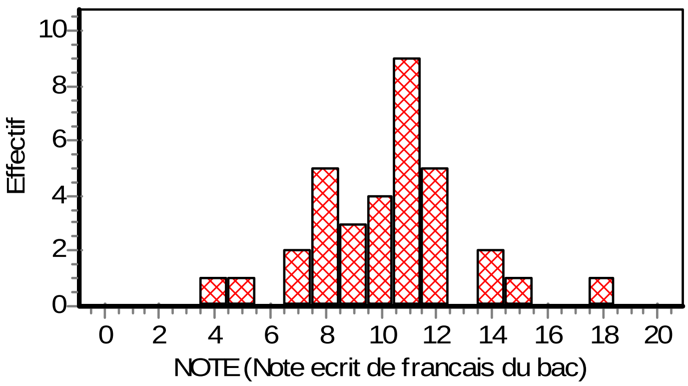
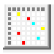


Figure 1 : Distribution des notes (histogramme)

- cette distribution apparaît relativement symétrique,
- on peut discuter du caractère relativement atypique de la valeur 18,
- la distribution se rapproche d'une distribution normale.

Pour voir le tableau des effectifs et des pourcentages :

```
Au-dessus du graphique précédent, cliquer sur
```


```
ou
Menu Statistiques
– Distribution
- Distribution des valeurs observées (effectifs et %)
```

Tableau 3 : Indices de tendance centrale

NOTE	4	5	7	8	9	10	11	12	14	15	18	Total
n	1	1	2	5	3	4	9	5	2	1	1	34
%	2.9%	2.9%	5.9%	14.7%	8.8%	11.8%	26.5%	14.7%	5.9%	2.9%	2.9%	100.0%

Synonymes

- distribution normale / distribution de Gauss /distribution de Laplace-Gauss / courbe en cloche / bell curve

A partir de quel moment qualifier une distribution de dissymétrique ? Si l'on ne se satisfait pas d'une analyse visuelle, il existe des indices de symétrie/dissymétrie basés sur la comparaison des indices de tendance centrale (cf. Tableau 4). De même comment qualifier une ou plusieurs valeurs d'atypiques ? A partir de quel moment peut-on dire que la distribution se rapproche d'une distribution normale ? Ces questions sont légitimes mais la statistique n'y apporte pas de réponses simples.

Tendance centrale ?

Dans l'échantillon?

L'indice de tendance centrale le plus connu est la moyenne. Toutefois, il est souvent intéressant de connaître la valeur d'un autre indice : la médiane. Enfin, on peut discuter du rattachement, ou non, d'un troisième indice à cette catégorie : le mode.

```
Menu Statistiques
- Tendance centrale
- Tous indices de tendance centrale
```

Tableau 4 : Indices de tendance centrale

Moy	10.3
Med	11.0
Mod	11.0

On constate que :
- la moyenne de ces notes est légèrement supérieure à 10 sur 20 (moy = 10.3),
- environ 50% des étudiants ont une note inférieure ou égale à 11 et donc environ 50% ont une note égale ou supérieure à 11 (med = 11),
- la note la plus fréquente est 11 (mod = 11).

Quel indice choisir ? Lorsqu'on a le choix entre plusieurs indices, la règle est… de ne pas choisir ! Chacun apporte un éclairage particulier sur les données. Ainsi, lorsque l'on s'intéresse à la distribution des salaires en France (cf. INSEE, 2012), il est informatif de savoir à la fois que le salaire moyen est autour de 2 100 euros net par mois, mais aussi qu'un salarié sur deux gagne moins de 1 700 euros (médiane) et enfin que le salaire le plus fréquent est proche de 1 200 euros (mode).

On se demandera si les indices de tendance centrale (moyenne, médiane, mode) sont proches ou, au contraire, divergent beaucoup. Dans ce dernier cas il faudra s'interroger sur les raisons de cette divergence : dissymétrie de la distribution ? valeur(s) atypique(s) ?
Les valeurs de ces trois indices sont confondues en particulier, mais pas uniquement, lorsque la distribution est symétrique et n'a qu'un mode (cf. l'exemple de la distribution normale). Si ces indices divergent, comme dans le cas des salaires, un seul indice ne peut pas suffire à résumer la tendance centrale.

Référence : Eisenhauer (2002)

Attention : Ce que l'on désigne par le *mode* ou la *valeur modale* est la valeur de la variable (ici 11) et non pas le nombre de fois où cette valeur a été observée (ici 9).

Dans notre exemple, les valeurs des trois indices diffèrent peu. Cela traduit le fait que la distribution est très peu dissymétrique.

Dans la population ?

L'échantillon analysé jusqu'à présent comporte 34 étudiants en 3ème année de psychologie à l'université Paris Descartes (Paris 5). Nous allons vérifier dans quelle mesure il est possible d'estimer certaines caractéristiques de la population (inconnue) à partir de l'échantillon (connu).

Il arrive fréquemment que l'on analyse, du point de vue d'une variable, un groupe limité d'individus dans la perspective de connaître, du point de vue de cette même variable, une population plus vaste dont fait partie ce groupe. C'est le cas ici : le groupe de 34 étudiants est un *échantillon* de la *population* des étudiants de cette année. On souhaite avoir une idée de la note moyenne de ces étudiants. Deux solutions : analyser la totalité des questionnaires et on connaîtra cette moyenne avec certitude ou, si on en n'a, ni le temps, ni le courage, ni les financements nécessaires, on pourra avoir recours aux procédures statistiques qui nous donneront une estimation de cette répartition.

La moyenne des notes observées sur l'échantillon est égale à 10.3. Qu'en est-il dans la population parente ? Autrement dit, quelle est la valeur de la *moyenne parente* ? Aucune procédure statistique ne permet de dire quelle est précisément cette valeur. On peut seulement calculer un intervalle de confiance, c'est-à-dire l'ensemble des valeurs possibles de cette moyenne parente.

Intervalle de confiance

```
Menu Statistiques
- Tendance centrale
- IC sur moyenne
```

Tableau 5 : Intervalle de confiance sur la moyenne parente

Moy	**10.3**
EtyC	2.75
ErreurTyp	0.47
Garantie	95%
Lim Inf.	9.34
Lim Sup.	11.30

La moyenne parente est comprise entre 9.3 et 11.3 (au seuil .05 / au niveau de confiance 95%).

Il importe de définir le plus précisément la population dont il est question. Ici, ce n'est pas la population des étudiants en psychologie, mais celle de tous les étudiants de psychologie…
… de cette année universitaire (1998-1999),
… de cette année d'études (3ème),
… de cette université (Paris Descartes).

Rédiger un compte-rendu de l'analyse

On s'intéresse aux notes obtenues à l'écrit de français du baccalauréat,
par les étudiants en 3ème année de psychologie à l'université Paris Descartes (Paris 5) en 1998-1999.
Ils ont été interrogés via un questionnaire.
On a constaté,
sur un échantillon de 34 de ces étudiants
que la moyenne de leurs notes est de 10.3.
Il semble que,
chez l'ensemble des étudiants interrogés,
la moyenne des notes
soit comprise entre 9.3 et 11.3
(*IC* 95% = [9.34 ; 11.3]).

Dispersion des valeurs?

L'analyse de la dispersion consiste à :
- repérer le minimum et le maximum (et éventuellement l'étendue),
- calculer des indices qui résument la dispersion des valeurs, soit autour de la moyenne, soit autour de la médiane.

Étendue (minimum et maximum)

```
Menu Statistiques
- Dispersion
- Étendue (Min, Max)
```

Tableau 6 : Minimum et maximum

Min	4.00
Max	18.00
Etendue	14.00

Pour ce groupe de 34 observations, les notes observées vont de 4 à 18.

Dispersion autour de la moyenne

Plusieurs indices, relativement équivalents, peuvent être utilisés pour résumer la dispersion des valeurs par rapport à la moyenne : l'écart absolu moyen, l'écart-type et l'écart-type corrigé. Tous ces indices sont des moyennes des écarts à la moyenne. L'écart absolu moyen (Eam) est le plus naturel : c'est la moyenne des valeurs absolues des écarts. L'écart-type (Ety) est également une moyenne des écarts[1]. C'est l'indice le plus souvent fourni. Notons toutefois qu'il présente l'inconvénient d'être sensible aux valeurs atypiques.

```
Menu Statistiques
- Dispersion
- Variances et écarts-type
```

Tableau 7 : Indices de dispersion autour de la moyenne

Variables	**Var**	**Ety**	**VarC**	**EtyC**
NOTE	7.33	2.71	7.55	2.75

On s'intéresse ici uniquement aux écarts-type (les variances sont des intermédiaires de calcul pour l'écart-type).[2]

Pour ce groupe de 34 étudiants de 3ème année de psychologie de l'université Paris Descartes (Paris 5), les notes à l'écrit de français s'écartent en moyenne de presque 3 points par rapport à la moyenne (ety = 2.71 et etyC =2.75).

Une variante de l'écart-type (Ety) est l'écart-type corrigé (EtyC) qui présente l'avantage, si l'on se place dans une perspective inférentielle, d'être un meilleur estimateur de l'écart-type de la population. Notons qu'il est parfois difficile de savoir, pour un utilisateur d'un logiciel statistique, si l'écart-type reporté est la version corrigée (EtyC) ou la version non corrigée (Ety).

Synonymes

- écart-type (non corrigé) / standard deviation / population standard deviation (SD)
- écart-type corrigé / sample standard deviation (s)

[1] L'écart-type doit être interprété comme une moyenne, même si c'est une moyenne *quadratique* (racine carrée de la moyenne des carrés des écarts à la moyenne).
[2] *Ety = racine(Var)*

Répartition par quartiles

Les quantiles (terme générique) sont des classes d'effectifs égaux. Selon l'effectif total du groupe observé, on analysera la répartition des valeurs selon des classes plus ou moins nombreuses. Dans le cas de petits effectifs comme ici, on calculera des quartiles (4 classes). Dans le cas d'effectifs plus importants, on calculera des déciles (10 classes) ou des centiles (100 classes).

```
Menu Statistiques
- Dispersion
- Boite à moustaches (médiane et quartile)
```

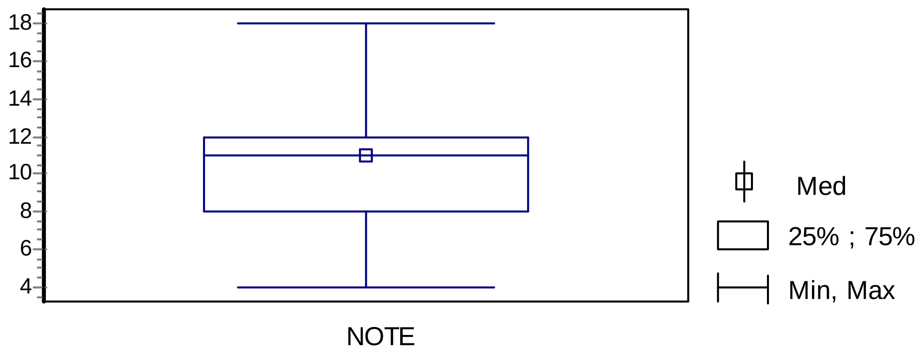


Figure 2 : Répartition des notes selon les quartiles (boîte à moustaches)

Pour voir le tableau des valeurs correspondant à ce graphique (min, max, médiane et quartiles):

```
Au-dessus du graphique précédent, cliquer sur
ou
Menu Statistiques
– Dispersion
- Quartiles
```

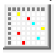


Tableau 8 : Indices de tendance centrale

Variables	Min	Q1	Med	Q3	Max
NOTE	4	8	11	12	18

Pour ce groupe de 34 étudiants en 3ème année de psychologie à l'université Paris Descartes (Paris 5), leurs notes à l'écrit de français varient de 4 à 18.
Elles se répartissent ainsi :
- environ 25% des notes[3] sont comprises entre 4 (min) et 8 (Q1),
- environ 50% des notes sont comprises entre 8 (Q1) et 12 (Q2),
- environ 25% des notes sont comprises entre 12 (Q3) et 18 (max).

Une autre option intéressante (non disponible dans la version actuelle de SES-Pegase) consiste à calculer des quintiles. On obtient une répartition en 5 classes (avec chacune 20% de l'effectif total) dont une classe centrale autour de la médiane et deux classes de chaque côté de celle-ci.

[3] *Ces pourcentages sont exacts seulement lorsque les quartiles ne correspondent pas à des valeurs observées.*

SITUER LES INDIVIDUS DANS LE GROUPE

Il existe plusieurs manières de positionner un individu par rapport au groupe auquel il appartient. Il est possible de :
- situer sa valeur par rapport à la moyenne du groupe,
- situer sa valeur parmi les différents quantiles (quartiles, déciles, centiles…),
- calculer la proportion de valeurs inférieures ou supérieures à sa valeur.

Écart à la moyenne du groupe

Lorsque l'on positionne la valeur observée pour un individu donné par rapport à la moyenne, on relève deux types d'informations :
- le sens de l'écart. La valeur de l'individu se situe-t-elle au-dessus ou en-dessous de la moyenne ?
- l'ampleur de l'écart. On peut alors s'appuyer sur deux types d'indices, un indice d'écart brut (l'écart à la moyenne du groupe) et un indice d'écart relatif (*l'écart centré-réduit* ou *score z*).

```
Menu Statistiques
– Écarts individuels à la moyenne
- Écarts bruts à la moyenne

Menu Statistiques
– Écarts individuels à la moyenne
- Écarts-centrés-réduits (scores z)
```

Les valeurs calculées ont été rassemblées dans le tableau suivant (cf. Tableau 9) :

Tableau 9 : écarts à la moyenne et scores z (extrait)

	NOTE	NOTE0	NOTEZ
SRP	9.00	-1.29	-0.48
EMA	11.00	0.71	0.26
SABX	8.00	-2.29	-0.85
EMIC	8.00	-2.29	-0.85
AMELIEM	8.00	-2.29	-0.85
CHEBLI	10.00	-0.29	-0.11
FUERTES	18.00	7.71	2.85
(...)	(...)	(...)	(...)
ZORRO	8.00	-2.29	-0.85
MELIS	10.00	-0.29	-0.11

Prenons l'exemple de l'étudiant dont le pseudo est Fuertes :
- sa note (18) est au-dessus de la moyenne du groupe (18 > 10.29),
- elle s'écarte donc de 7.71 points au-dessus de la moyenne du groupe (NOTE0 = 7.71),
- elle se situe à presque 3 écarts-type (NOTEZ = 2.85) au-dessus de la moyenne du groupe.

Prenons un autre exemple, celui de l'étudiant dont le pseudo est Zorro (cf. Tableau 9) :
- sa note (8) est au-dessous de la moyenne du groupe (8 < 10.29),
- elle s'écarte donc d'un peu plus de deux points au-dessous de la moyenne du groupe (NOTE0 = -2.29),
- elle se situe à presque un écart-type au-dessous de la moyenne du groupe type (NOTEZ = -0.85).
cf. Annexe pour le calcul de ces valeurs.

L'intérêt de l'écart centré-réduit ou score z est qu'il prend en compte la dispersion des valeurs du groupe. Si la moyenne des notes est 11, une note de 13, à 2 points au-dessus de la moyenne du groupe, n'a pas la même signification selon que les autres notes du groupe s'écartent en moyenne de 1 point de la moyenne (faible dispersion) ou de 6 points (forte dispersion). Le score z sera plus élevé dans le premier cas (13 fait partie des meilleures notes) que dans le second (13 est, relativement aux autres notes, proche de la moyenne).

Position dans les quantiles

Considérons à nouveau les valeurs de la dispersion autour de la médiane. Elles indiquent les valeurs du 1er quartile (Q1=8), de la médiane (med = Q2 = 11) et du 3ème quartile (Q3=12).

Reprenons l'exemple de l'étudiant dont le pseudo est Fuertes et dont la note est 18. Cet étudiant se trouve donc dans le 4ème quartile, entre Q3 et le maximum, c'est-à-dire que sa note fait partie des 25% meilleures notes.
Autre exemple, considérons l'étudiant dont le pseudo est SRP (cf. Tableau 9) et dont la note est 9. Cet étudiant se trouve dans le deuxième quartile, entre Q1 et Q2 (la médiane).

Lorsque l'effectif du groupe est grand on pourra placer un score donné, non seulement par rapport aux quartiles, mais aussi par rapport à des quantiles plus fins (déciles ou centiles). Si l'on considère la distribution des salaires, il est certes intéressant de savoir qu'un salarié est au-dessus de la médiane ou dans le quatrième quartile (son salaire fait partie des 25% les plus élevés). Il est beaucoup plus intéressant de savoir qu'il se trouve dans le 10ème décile (son salaire fait partie des 10% les plus élevés) voire dans le 100ème centile (son salaire fait partie des 1% les plus élevés).

Pourcentages inférieur et supérieur

Dans le prolongement de ce qui précède (position par rapport aux quantiles) une autre alternative consiste à situer un individu en calculant les proportions des autres individus dont la note est inférieure ou supérieure à la sienne.
Prenons l'exemple de Zorro, dont la note est 8.

```
Menu Statistiques
- Situer une valeur dans le groupe
- Indiquer la valeur : 8 OK
Menu Statistiques
- Situer une valeur dans le groupe
- Répartition en % par rapport à la valeur
```

Tableau 10 : Position de la note de « Zorro » (8) dans la distribution

NOTE	n	%
< 8	4	12%
= 8	5	15%
> 8	25	74%

On constate que :
- 4 notes sur 34, soit 12% des notes, inférieures à sa note (cf. Tableau 10),
- 5 notes, dont la sienne, soit 15% des notes, identiques à sa note,
- 25 notes sur 34, soit 74% des notes, supérieures à sa note.

Rang percentile

Le rang percentile est le pourcentage des individus du groupe dont le score est inférieur ou égal à l'individu considéré. Ainsi calculons le rang percentile de Zorro dont la note est 8.

```
Menu Statistiques
- Situer une valeur dans le groupe
- Rang percentile
```

Tableau 11 : Rang percentile d'un individu ayant un score de 8

NOTE	n	Rang%
<= 8	9	26.50%

Dans ce groupe, 26.5% des individus (9/34) ont une note inférieure ou égale à celle de Zorro (8). Son rang percentile est 26.5%.
Le rang percentile de l'individu est d'autant plus élevé que son score est élevé. L'individu dont le rang percentile est 100% est celui, ou parmi ceux, qui a/ont le meilleur score du groupe.
Synonyme :

- rang percentile / percentile rank

SITUER LE GROUPE PAR RAPPORT À UNE VALEUR DE RÉFÉRENCE

Il existe souvent, pour une variable donnée, une valeur de référence : 10/20 pour des notes sur 20 ; 100 pour des QI… Cette valeur est à indiquer par l'utilisateur. S'agissant ici de notes, on prendra l'exemple d'une valeur de référence égale à 10. On cherchera alors à situer la moyenne du groupe par rapport à cette valeur de référence.

```
Menu Statistiques
- Écart du groupe à une valeur de référence,
- Choix de la valeur de référence,
Indiquer 10 OK
```

Sens de l'écart ?

Dans l'échantillon ?

```
Menu Statistiques
- Écart du groupe à une valeur de référence,
Sens de l'écart à la valeur de référence
```

Tableau 12 : Moyenne observée et valeur de référence

M_obs	10.3
Ref	10.0

On constate (cf. Tableau 12) que la moyenne observée dans l'échantillon (*moy* = 10.3) est au-dessus de la valeur de référence (*Réf* = 10).

Dans la population ?

Pour ce groupe de 34 observations, la moyenne des notes est de 10.3, donc supérieure à la valeur de référence (10/20). Que peut-on dire de la position de la moyenne de la population (moyenne parente) par rapport à cette valeur de référence ? Le *test t de Student de comparaison d'une moyenne à une valeur de référence* permet de répondre à cette question

Test t de Student de comparaison d'une moyenne à une valeur de référence

```
Au-dessus du tableau précédent, cliquer sur Inférer
ou
Menu Statistiques
- Écart du groupe à une valeur de référence,
- t de Student pour écart à la valeur de référence.
```

Tableau 13 : Test t de Student de comparaison d'une moyenne à une valeur de référence

Moy	10.3
Réf	10
t	0.62
ddl	33
p	.5367

Ce test *t* de Student teste l'hypothèse selon laquelle la moyenne parente est supérieure à la valeur de référence (10). Si $p < .05$ (ce qui n'est pas le cas ici) le test est déclaré, selon la formule usuelle, "significatif" et on en conclut que la moyenne parente est, comme dans l'échantillon, supérieure à la valeur de référence.
Si $p > .05$, le résultat du test est déclaré "non significatif" et on ne peut pas conclure que la moyenne parente est supérieure, ni inférieure, ni égale à cette valeur de référence.

Attention :
- Ce test non significatif ne permet pas de conclure que la moyenne parente est égale à 10. Dire « on ne peut pas conclure qu'elle est différente de 10 » n'est pas équivalent à dire « elle est égale à 10 ».
- Pour le cas où le test aurait été significatif, cela aurait permis de conclure que la moyenne parente est supérieure à 10/20. Mais cela n'aurait pas permis de savoir si elle s'écarte peu ou beaucoup de 10/20. L'expression « différence significative » souvent utilisée à la suite d'un test significatif, est à éviter car elle peut suggérer au lecteur que la différence entre la moyenne parente et la valeur de référence est importante.

Ampleur de l'écart ?

Dans l'échantillon ?

Écart brut

```
Menu Statistiques
- Écart du groupe à une valeur de référence,
- Écart brut à la valeur de référence.
```

Tableau 14 : écart brut entre moyenne et valeur de référence

Moy	10.29
Ref	10.00
d	+0.29

La moyenne du groupe (10.29) est à 0.29 point au-dessus de la valeur de référence (10/20). Cet écart est faible, car inférieur à 1 point (cf. présentation des données).

Écart calibré

L'intérêt de l'écart calibré est de prendre en compte la dispersion des valeurs (à travers l'écart-type de ces valeurs). Un même écart brut conduira à un écart calibré d'autant plus élevé que la dispersion est faible.

```
Menu Statistiques
- Écart du groupe à une valeur de référence,
- Écart calibré à la valeur de référence
```

Tableau 15 : écart calibré entre moyenne et valeur de référence

Moy	10.29
Ref	10.00
Ety	2.71
ECM	0.11

L'écart calibré (*ECM=d/ety*) de la moyenne à la valeur de référence est égal à 0.11. On conviendra que cet écart est faible, car inférieur à 0.40 (il aurait été qualifié de fort s'il avait été supérieur à 0.87).

Pour les notes à l'écrit de français
de ce groupe de 34 étudiants en 3$^{\text{ème}}$ année de psychologie à l'université Paris Descartes (Paris 5),
la moyenne (moy = 10.3) est supérieure à la moyenne de référence (10/20).
Toutefois l'écart apparaît faible,
que l'on considère l'écart brut ($|d|$ = 0.29 < 1 point)
ou l'écart calibré ($|EC|$ = 0.11 < 0.30).

Dans la population ?

Intervalle de confiance sur l'écart à la valeur de référence

```
Menu Statistiques
- Écart du groupe à une valeur de référence,
- IC pour écart à la valeur de référence.
```

Tableau 16 : Intervalle de confiance sur l'écart entre moyenne et valeur de référence

	Moy	Ref	d_obs	p	Erreur-ty	h_alpha	IC_inf	IC_sup
NOTE	10.3	10	0.29	0.05	0.47	0.96	-0.66	+1.25

Autre procédure :

À partir de la fenêtre précédente, cliquer sur inférer
Sélectionner Intervalle de confiance sur l'écart à la valeur de référence
Cliquer sur OK

Cet intervalle de confiance ([-0.66 ; +1.25]) nous indique que :
1. l'écart entre la moyenne parente et la valeur de référence est peut-être positif mais peut-être aussi nul (0 est une valeur comprise dans l'intervalle) et enfin… peut-être négatif puisqu'il est compris entre -0.66 et +1.25. Ce résultat est cohérent, nécessairement, avec le résultat (non significatif) du test *t* de Student. Tous deux nous conduisent à la conclusion que… il est impossible de conclure quant à la position de la moyenne parente par rapport à la valeur de référence.
2. l'écart de la moyenne parente à la valeur de référence est compris entre (entre -0.66 et +1.25). Cet écart parent pouvant donc être supérieur à 1, il n'est pas possible de conclure à un écart faible.

Rédiger le compte rendu de l'analyse

Nous avons constaté,
sur un échantillon de 34 étudiants en 3ème année de psychologie à l'université Paris Descartes (Paris 5),
que la moyenne de leurs notes à l'écrit du baccalauréat de français (*moy* = 10.3)
est supérieure à la moyenne de référence (10/20).
Toutefois, il semble que,
pour l'ensemble des étudiants dont a été extrait cet échantillon,
il n'est pas possible de conclure que la moyenne de leurs notes est supérieure à 10/20
($t_{[33]}$ = 0,62 , p = .54, > .05).
Celle-ci peut être, par rapport à 10/20, inférieure de 0.66 points ou supérieure de 1.25 points
(*IC* 95% = [-0.66 ; +1.25]).

ANNEXE – PROCÉDURES DE CALCUL

Soit le tableau suivant de 10 valeurs numériques quelconques (cf. colonne X) obtenues par 10 personnes.
On note *n* le nombre de valeurs.

Tableau 17 : Tableau de valeurs fictives

	X	X-moy	|X-moy|	(X-moy)²
i01	5	+2	2	4
i02	0	-3	3	9
i03	5	+2	2	4
i04	1	-2	2	4
i05	0	-3	3	9
i06	3	0	0	0
i07	0	-3	3	9
i08	5	+2	2	4
i09	6	+3	3	9
i10	5	+2	2	4

Calcul des indices de tendance centrale

Moyenne

Moy = 3
La somme des notes est 30. D'où une moyenne égale à 3 (30/n = 30/10 = 3)

Médiane

Méd = 4
Si l'on ordonne les 10 notes : 0, 0, 0, 1, 3, 5, 5, 5, 5, 6, on constate que la médiane est comprise entre 3 et 5. On peut prendre comme valeur de la médiane, le centre de cet intervalle, soit 4. La valeur 4 partage le groupe en deux classes de même effectif : 50% des notes sont inférieures à 4 et 50% sont supérieures à 4.

Mode

Mod = 5.
La valeur la plus fréquente est 5, observée 4 fois. C'est le mode ou valeur modale.

Calcul des degrés de liberté (ddl)

dl = 10 – 1 = 9

Calcul des indices de dispersion

Écart absolu moyen

Eam = 2.2
C'est la moyenne des valeurs absolues des écarts à la moyenne (cf. colonne |X-moy|).
La somme est égale à 22, d'où une moyenne de 2.2 (22 / 10).

Variance

Var = 5.6
Le calcul de la variance passe par le calcul de la somme des carrés des écarts à la moyenne (cf. colonne (X-moy)²).
On la note SC (Somme des Carrés).
On a SC = 4 + 9+… + 4 = 56.
Var est égale à SC / n = 56 / 10 = 5.6

Écart-type

Ety = 2.37
C'est la racine carrée de la variance.
Ety = racine(Var) = 2.37

Variance corrigée

VarC = 6.22
VarC = SC / ddl = 56 / 9 = 6.22

Écart-type corrigé

EtyC = 2.49
EtyC = racine(VarC) = racine(6.22) = 2.49

Calcul des écarts d'un individu à la moyenne du groupe

Pour i09, la valeur observée est égale à 6. Cette valeur est :
- au-dessus de la moyenne (6 > 3)
- à 3 points au-dessus de la moyenne (6 - 3)
- à 1.27 écart-type au-dessus de la moyenne ((6 – 3) / 2.37 = 1.27)

Pour i04, la valeur observée est égale à 1. Cette valeur est :
- au-dessous de la moyenne (1 < 3)
- à 2 points en-dessous de la moyenne (1 - 3 = -2)
- à 0.85 écart-type en-dessous de la moyenne (-2 / 2.37 = -0.85)

RÉFÉRENCES

APA. (2010). *Publication Manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Corroyer, D., & Rouanet, H. (1994). Sur l'importance des effets et ses indicateurs dans l'analyse statistique des données. *Année Psychologique, 94*(4), 607-624.

Corroyer, D., & Wolff, M. (2003). *L'Analyse Statistique des Données en Psychologie; Concepts et Méthodes de base.* Paris: Armand Colin (Cursus).

Eisenhauer (2002). Symmetric or Skewed, *The College Mathematics Journal*, 33, 1

Rouanet, H., Bernard, J.-M., & Le Roux, B. (1990). *Statistique en Sciences Humaines. Analyse Inductive des Données*. Paris: Dunod.

Rouanet, H., Lecoutre, M. P., Bert, M. C., Lecoutre, B., & Bernard, J.-M. (1991). *L'inférence statistique dans la démarche du chercheur*. Berne: Peter Lang.

LISTE DES TABLEAUX

LISTE DES FIGURES

SOMMAIRE