

UN GROUPE D'INDIVIDUS DÉCRIT PAR UNE VARIABLE QUANTITATIVE

SITUER LE GROUPE PAR RAPPORT À UNE POPULATION DE RÉFÉRENCE

Mots clés : écarts à la moyenne ; écarts-réduit ; score Z ; distribution normale ; approximation normale ; situer un groupe par rapport à une distribution de référence.

LE DOSSIER SPATIAL¹

Le groupe observé

On dispose des données recueillies sur un groupe de 16 garçons de troisième. Ils ont passé un test spatial. La moyenne de leurs scores (nombre de bonnes réponses à 48 questions) est 30.

La population de référence

Lors d'une enquête réalisée en 1964 sur 5154 garçons de Troisième on a relevé les à ce même test spatial. Il s'agissait de constituer une "population de référence" sur les capacités spatiales des garçons de Troisième.

Tableau 1 : Distribution des scores au test spatial

x !	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
n !	1	2	1	1	2	2	5	12	10	19	29	42	58	65	80	110	162	139
x !	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33		
n !	152	185	193	213	270	333	315	310	263	280	312	239	224	215	181	167		
x !	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48			
n !	153	83	93	68	74	42	20	12	8	4	3	1	0	1	0			

Question

On se demande si ce groupe est atypique de la population de référence.

Afin de pouvoir situer le groupe observé par rapport à la population de référence, il faut commencer par décrire précisément cette population de référence.

Les données de la population se trouvent dans le fichier SPATIAL.SES

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

¹ Source : Bacher et Nguyen-Xuan, BINOP, 1967.

ANALYSER LA POPULATION DE RÉFÉRENCE

Menu Nouvelle analyse
Sélectionner SCORES comme "Variable(s) à analyser"

Forme de la distribution ?

Menu Statistiques
Histogramme

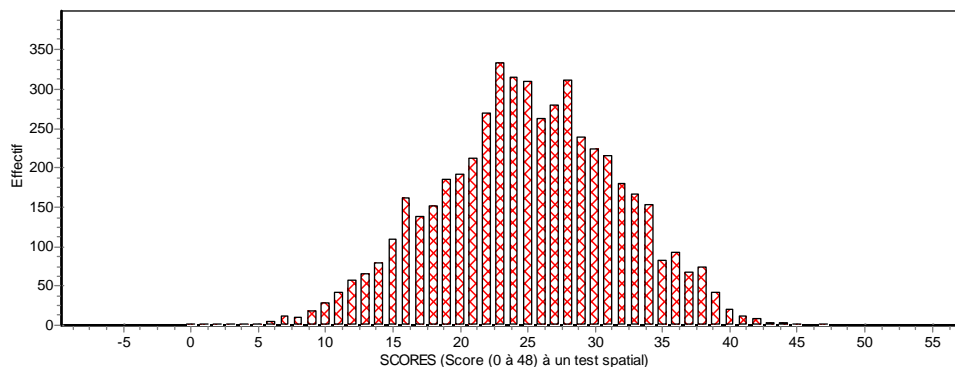


Figure 1: Distribution des scores au test spatial de la population de référence

Cette distribution:

- a une forme « en cloche » qui est proche de la forme d'une distribution normale,
- est quasi unimodale,
- est quasi symétrique.

Tendance centrale ?

Menu Statistiques
- Tendance centrale
- Tous indices de tendance centrale

Tableau 2 : Moyenne, médiane et mode de la population de référence

	SCORES
Moy	25
Med	25
Mod	23

On constate que la moyenne et la médiane sont identiques (25).² Le mode (23) est un peu différente de ces deux indices.

La proximité des indices de tendance centrale reflète la symétrie de la distribution constatée sur l'histogramme.

Dispersion ?

Menu Statistiques
- Dispersion
- Variances et écarts-type

Tableau 3 : Variances et écarts-type (non corrigés et corrigés)

	Var	Ety	VarC	EtyC
SCORES	46.1	6.79	46.1	6.79

Les versions non corrigées et corrigées diffèrent ici très peu (seulement à la troisième décimale pour les écarts-type). Ceci est dû à l'effectif très élevé (n= 5154).

² En fait la moyenne est égale, plus précisément, à 25.05

Situer un individu par rapport à la population

Où se trouve un individu ayant obtenu, par exemple, un score de 30 au test spatial ?

Menu Statistiques
- Situer une valeur dans le groupe
- Indiquer la valeur : 30

Il y a plusieurs moyens de répondre à cette question :

- en calculant le nombre d'individus du groupe ayant eu un score inférieur/égal/Supérieur au score de cet individu,
- en calculant l'écart de ce score à la moyenne des scores du groupe,
- en calculant l'écart centré-réduit (score z) de ce score à la moyenne des scores du groupe.

Menu Statistique
- Situer une valeur dans le groupe
- Répartition en % du groupe par rapport à la valeur

Tableau 4 : Répartition en % du groupe par rapport à la valeur 30

SCORES	n	%
< 30	3805	73.80%
30	224	4.30%
> 30	1125	21.80%

Ce tableau nous indique que, dans cette population de référence :

- un peu moins des trois-quarts des scores des individus sont inférieurs à 30,
- environ 5% (4.3%) sont égaux à 30,
- un peu plus de 20% (21.8%) sont supérieurs à 30.

Menu Statistique
- Situer une valeur dans le groupe
- Écart de la valeur à la moyenne

Tableau 5 : Écart de la valeur 30 à la moyenne du groupe

	SCORES
moy	25
x	30
x-moy	4.95

La valeur 30 est à presque 5 points (4.95) au dessus de la moyenne.

Menu Statistique
- Situer une valeur dans le groupe
- Écart centré-réduit (score z) à la moyenne

Tableau 6 : Écart centré-réduit (score z) de la valeur 30 à la moyenne du groupe

	SCORES
Moy	25
Ety	6.79
x	30
z	0.73

La valeur 30 est à moins d'un écart-type ($z = 0.73$) au dessus de la moyenne³.

Approximation par une distribution normale

Nous avons vu que la distribution des scores est proche d'une distribution normale. Par ailleurs la moyenne des scores est égale à 25 (plus précisément 25.05) et l'écart-type est égal à 6.8.

Peut-on approximer cette distribution par une distribution normale de moyenne 25 et d'écart-type 6.79 ?

Dit autrement, que se passerait-il si l'on considérait dorénavant, non pas la distribution réelle des scores, mais l'approximation normale de ces scores?

³ $z = (30-25.05)/6.79$

Par exemple, la position d'un score de 30 dans cette distribution normale serait-elle très différente de celle obtenue ci-dessus ?

Menu Outils
- Loi Normale
Indiquer la moyenne (25.05), l'écart-type (6.79) et la valeur (30)

Le logiciel calcule le score z (0.73). Il nous indique que :

- environ 77% (76.7%) des valeurs sont inférieures à 30,
- environ 23% (23.3%) sont supérieures à 30.⁴

Ces valeurs (77% et 23%) issues d'une approximation normale de la distribution sont à comparer aux proportions exactes calculées précédemment (74% et 22%).

On constate que cet ajustement n'est pas parfait.

SITUER LE GROUPE PAR RAPPORT À LA POPULATION DE RÉFÉRENCE

Rappelons que l'on dispose des données recueillies sur une groupe de 16 garçons de troisième et on se demande si ce groupe est atypique d'une population de référence.

Pour répondre à cette question, on va situer ce groupe particulier, de 16 personnes, par rapport à tous les échantillons de 16 personnes que l'on aurait pu constituer à partir de la population de référence.

La distribution d'échantillonnage de la moyenne (DEM)

La théorie nous indique que si on calcule les moyennes des scores de tous les échantillons possibles, la distribution de toutes ces moyennes :

- a pour moyenne : la moyenne de la distribution de référence (25.05)
- a approximativement pour écart-type : l'écart-type de la distribution de référence divisé par racine(16), soit $6.79 / 4 = 1.7$.
- a une forme proche d'une distribution normale, dès lors que la distribution de référence n'est pas trop dissymétrique⁵.

La distribution de toutes les moyennes de tous les échantillons possibles est appelée *Distribution d'échantillonnage de la Moyenne*, en bref DEM.

Situer le groupe dans la DEM

Pour situer le groupe par rapport à la population de référence, on va situer sa moyenne par rapport à toutes les moyennes de tous les échantillons, c'est-à-dire dans la DEM.

Menu Outils
- Loi Normale
Indiquer la moyenne (25.05), l'écart-type (1.7) et la moyenne de l'échantillon (30)

Le logiciel calcule le score z (2.91). Cela signifie que ce groupe de 16 individus est à presque 3 écarts-type au-dessus de la moyenne.

Il calcule la proportion des échantillons qui ont une moyenne supérieure à 30. Cette proportion est de 0.0018 (0.18%).

On considère que l'échantillon est atypique si cette proportion est faible. Par convention on considère que cette proportion est faible si elle est inférieure à 2.5%.⁶

Ici on pourra donc conclure que cet échantillon est atypique de la population de référence ($p = 0.18\% < 2.5\%$).

⁴ Dans toute distribution continue la proportion (ou probabilité) des valeurs égales à une certaine valeur est 0.

⁵ Si la distribution de référence est proche d'une distribution normale, la distribution des moyennes sera "encore plus proche" d'une distribution normale.

⁶ On prend 2.5% comme seuil repère conventionnel car il s'agit d'un proportion unilatérale. Dans le cas d'une proportion (ou probabilité) bilatérale, la valeur repère conventionnelle est 5% ($2 * 2.5\%$).

RÉFÉRENCES

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

Liste des Tableaux

Tableau 1 : Distribution des scores au test spatial	1
Tableau 2 : Moyenne, médiane et mode de la population de référence.....	2
Tableau 3 : Variances et écarts-type (non corrigés et corrigés)	2
Tableau 4 : Répartition en % du groupe par rapport à la valeur 30.....	3
Tableau 5 : Écart de la valeur 30 à la moyenne du groupe	3
Tableau 6 : Écart centré-réduit (score z) de la valeur 30 à la moyenne du groupe	3

Liste des Figures

Figure 1: Distribution des scores au test spatial de la population de référence	2
---	---

SOMMAIRE

Dossier SPATIAL	1
<i>Le groupe observé.....</i>	<i>1</i>
<i>La population de référence.....</i>	<i>1</i>
<i>Question</i>	<i>1</i>
Analyser la population de référence	2
<i>Forme de la distribution ?.....</i>	<i>2</i>
<i>Tendance centrale ?</i>	<i>2</i>
<i>Dispersion ?.....</i>	<i>2</i>
<i>Situer un individu par rapport à la population.....</i>	<i>3</i>
<i>Approximation par une distribution normale</i>	<i>3</i>
Situer le groupe par rapport à la population de référence	4
<i>La distribution d'échantillonnage de la moyenne (DEM)</i>	<i>4</i>
<i>Situer le groupe dans la DEM</i>	<i>4</i>
Références	5
Liste des tableaux.....	5
Liste des figures.....	5