



MIASS 231
Mathématiques (appliquées aux sciences sociales) 3
© El Hadj Touré, 2022

DIX EXERCICES RÉCAPITULATIFS
Leçon 5. Mesures de variation
(Solutionnaire)

1) Comparez, en termes de différence et de similitude, les paires de concepts ci-dessous :

a) Mesures de tendance centrale et mesures de dispersion

Les mesures de tendance centrale et les mesures de variation relèvent toutes les deux des **statistiques descriptives** puisqu'elles aident à décrire les données d'échantillon d'une variable par une seule valeur, un seul résumé quantitatif.

Toutefois, si les mesures de tendance centrale révèlent le score le plus commun dans une distribution afin de décrire la **représentativité d'un phénomène**, les mesures de variation, quant à elles, cherchent à mettre en évidence comment les scores sont éparpillés ou dispersés les uns par rapport aux autres et autour de la moyenne afin de connaître la **variabilité d'un phénomène**.

b) Étendue et intervalle interquartile

L'étendue et l'intervalle interquartile relèvent des **mesures de variation**, lesquelles cherchent à mettre en évidence comment les scores sont éparpillés ou dispersés les uns par rapport aux autres et autour de la moyenne afin de connaître la variabilité d'un phénomène. Elles ont ceci de commun qu'elles reposent seulement sur deux scores : elles n'apportent donc aucune information quant aux autres scores.

Toutefois, si l'étendue subit l'**effet des scores extrêmes** en ce qu'elle implique les valeurs maximale et minimale dans son calcul, ce n'est pas le cas de l'intervalle interquartile qui dépend plutôt des quartiles 1 et 3.

c) Variance et écart-type

L'écart-type et la variance relèvent des mesures de variation qui **exploitent l'ensemble de l'information** contenue dans les scores d'une distribution.

Toutefois, l'écart-type est plus approprié pour décrire la variabilité d'une distribution en raison de sa proximité avec les scores originaux. Par contre, la variance est surtout réservée aux statistiques inférentielles dont l'analyse de variance.

d) Asymétrie et kurtose

L'asymétrie et la kurtose constituent des mesures de la **forme d'une distribution**. Elles réfèrent aux moments d'ordre 3 et 4 dans la mesure de la variation d'une variable (sachant que la moyenne et la variance correspondent aux moments 1 et 2).

Toutefois, l'asymétrie illustre le degré d'étirement d'une distribution vers la gauche ou la droite (négative ou positive ou symétrique), alors que la kurtose révèle le degré d'aplatissement d'une distribution (aplatie, moyenne ou pointue).

2) Résolvez les deux problèmes pratiques en répondant aux questions :

2.1. Vous êtes recruté(e) par une compagnie aérienne, qui exploite des avions de 100 places, qui prend souvent plus de réservations qu'elle ne peut en accommoder, à cause des clients qui ne se présentent pas. Le problème est qu'il peut arriver que l'on dédommage des clients à qui on refuse l'embarquement lorsqu'un avion est rempli à pleine capacité. Pour faire face à ce dilemme, on vous demande en tant qu'expert statisticien, de déterminer le nombre idéal de réservations.

- a) Votre analyse va-t-elle porter sur la variation des pourcentages des clients qui se présentent ? Ou va-t-elle porter sur la tendance centrale des pourcentages des clients qui se présentent ? Justifiez votre point de vue.

Je chercherais à connaître le pourcentage de passagers ayant réellement embarqué pour tel vol, tel autre vol, encore et encore. Or, ces données varient selon l'heure, la saison, la nature du vol, etc. Les mesures de variation sont donc plus appropriées que les mesures de tendance centrale pour déterminer supposément le nombre idéal de réservations. Mon analyse reposerait sur les mesures de variation qui informent mieux sur la variation des pourcentages de personnes qui se présentent.

- b) Expliquez comment vous allez vous y prendre.

Concrètement, après avoir obtenu les **données** concernant les pourcentages de personnes qui se présentent pour une pluralité de vols, je calcule ensuite la **moyenne** puis l'**écart-type** de la distribution. Enfin, je calcule les **scores-z** pour chaque score brut (pourcentage) et ainsi, à l'aide de la courbe normale standardisée, je pourrais avoir une idée de la proportion de passagers comprise entre deux scores bruts (pourcentages). Par exemple, en appliquant la loi normale, je sais que 95% des passagers se situeront environ à plus ou moins 2 écarts-types de la moyenne.

NB : Les mesures de variation apportent souvent plus d'information que les mesures de tendance centrale. Remarquez que les mesures de variation (variance, écart-type, score-z) comprennent la moyenne 😊

2.2. Un chercheur s'intéresse aux inégalités de revenu entre les hommes et les femmes. Chacune des deux distributions n'étant pas symétrique et unimodale, il décide de standardiser les revenus des femmes et les revenus des hommes en utilisant les scores-Z dans l'espoir de normaliser les distributions afin de pouvoir les approximer par la loi normale et utiliser des techniques statistiques avancées.

a) Peut-il vraiment obtenir une distribution normale par le biais de ladite standardisation?

Il ne peut en aucun cas par ce biais obtenir une distribution normale. Les scores- Z_i permettent simplement de centrer (sur la moyenne) et de réduire (par rapport à l'écart-type) les scores X_i comme les revenus. Ils ne peuvent donc pas transformer une distribution asymétrique en une distribution normale. Par contre, l'avantage des scores- Z réside dans le fait qu'ils ramènent les scores originaux sur une échelle comprise entre -3 et 3 et ainsi offre une possibilité de comparaison avec la loi normale centrée réduite, pour des distributions différentes par exemple.

NB : Le centrage et la réduction des scores originaux X_i à l'aide des scores- Z_i ne transforment nullement une distribution « anormale » en une distribution normale. La standardisation n'implique pas la normalisation en conséquence. Toutefois, notez que l'utilisation de la loi normale pour résoudre des problèmes pratiques suppose que la variable d'intérêt suive une distribution normale et que les scores de la variable soient centrés et réduits à l'aide des scores- Z_i 😊

b) Y a-t-il une autre procédure statistique permettant de normaliser les deux distributions?

Oui, pour normaliser une distribution, les statisticiens ou analystes ont tendance à procéder à diverses **transformations**, dont le logarithme des variables à analyser. Ces transformations peuvent ainsi changer une distribution asymétrique en une distribution normale grâce à laquelle il est possible de mener des analyses reposant sur le postulat de la normalité, soit les statistiques inférentielles paramétriques (que nous aborderons dans la deuxième partie du cours).

3) Répondez aux questions objectives de compréhension ci-dessous :

3.1. Pour vérifier l'état général des bébés, un pédiatre-chercheur mesure le poids de 30 bébés à la naissance dans un grand hôpital. D'une part, la mesure en kilogrammes donne une moyenne de 3,5 kg et un écart-type de 0,5 kg. D'autre part, la mesure en grammes donne une moyenne de 3 500 g et un écart-type de 500 g.

a) Laquelle des deux distributions démontre une plus grande variabilité : celle qui est mesurée en kg ou en g?

À priori on peut croire que la distribution mesurée en g est plus disparate en raison de la grandeur des mesures. Mais, au fond, les deux distributions (en g comme en kg) démontrent une variabilité égale.

b) Justifiez, voire prouvez votre réponse.

Pour justifier mon point de vue, je me réfère au coefficient de variation : $s / X * 100$

Pour la distribution en kg, $CV = 0,5/3,5 * 100 = 14,3\%$

Pour la distribution en g, $CV = 500/3500 * 100 = 14,3\%$

NB : Par conséquent, pour comparer la variabilité de deux distributions dont l'unité de mesure est différente, le coefficient de variation s'avère approprié 😊

3.2. Le responsable de l'admission au sein d'un programme contingenté en psychologie veut classer et sélectionner des candidats issus de différents cégeps. Il souhaite utiliser les score-Z à cet effet.

- a) Peut-il se fier aux scores-Z aux fins du classement et de la sélection des candidats? Expliquez-vous!

Oui, il peut se fier aux scores-Z aux fins du classement et de la sélection puisque les scores-Z permettent de comparer les notes des étudiants sur une même échelle en prenant en compte les **différences artificielles** dans le système de notation et dans les épreuves tout en respectant, dans bien des cas, le classement originel des étudiants. Non seulement ils aident à comparer la position de la note d'un étudiant dans son groupe, mais aussi à comparer les positions des notes d'étudiants issus de groupes différents. C'est la raison pour laquelle les scores-Z ont été introduites dans le système éducatif québécois en 1977, en matière d'orientation des cégépiens.

- b) Quand on sait que les candidats proviennent de groupes aux caractéristiques différentes, peut-on vraiment affirmer qu'avec les scores-Z le rendement des élèves de tous les collèges devient comparable? Expliquez.

Il n'est pas juste de considérer qu'avec les scores-Z, le rendement des élèves de tous les Cégeps devient comparable à tout point de vue. Le principal problème est que les groupes n'ont pas souvent les mêmes caractéristiques et ne sont pas d'égale force. D'une part, les meilleurs étudiants sont pénalisés lorsqu'ils appartiennent à un groupe hétérogène où il y a beaucoup d'élèves moins forts (en fait, lorsque l'écart-type augmente, la cote-z tend à diminuer). D'autre part, la taille du groupe influe sur l'importance de la cote-z, les grandes classes tendant à augmenter les cotes-z des meilleurs étudiants. Si les scores-z permettent de résoudre les **différences artificielles**, ils ne permettent pas en revanche de résoudre les **différences structurelles** inhérentes aux groupes eux-mêmes.

NB : Comme toute mesure, les scores-z ne rendent pas compte de la complexité de la réalité scolaire (sociale) et ils doivent être interprétés avec beaucoup de prudence 😊

- 4) Vous vous intéressez à l'espérance de vie en années afin de décrire la variabilité de ce phénomène au sein de 10 pays du monde. Ci-dessous les données de l'échantillon (n=10) :

Individus (cas)	Espréance de vie
Allemagne	81
Argentine	77
Australie	83
Canada	82
Caméroum	59
Chine	77
Nigéria	55
Roumanie	75
Sénégal	68
Uruguay	78

a) Calculez et interprétez statistiquement l'étendue de la distribution.

$$E = X_{\max} - X_{\min}$$

$$E = 83 - 55 = 28$$

L'étendue est de 28. **Ce qui signifie que la distribution de l'espérance de vie au sein des 10 pays s'étend de 55 à 83 ans, soit un intervalle de 28.**

b) Sachant que l'espérance de vie moyenne est de 73,5 ans, calculez l'écart-type de l'espérance de vie. Interprétez statistiquement.

$$s = \sqrt{\frac{\sum (X - Moy)^2}{n - 1}}$$

$$s = \sqrt{\frac{(81 - 73,5)^2 + (77 - 73,5)^2 + (83 - 73,5)^2 + (82 - 73,5)^2 + (59 - 73,5)^2 + (77 - 73,5)^2 + (55 - 73,5)^2 + (75 - 73,5)^2 + (68 - 73,5)^2 + (78 - 73,5)^2}{10 - 1}}$$

$$s = \sqrt{\frac{845,5}{10 - 1}} = \sqrt{94,28} = 9,7$$

L'écart-type de l'échantillon équivaut à 9.7. **Autrement dit, environ deux tiers (68%) des 10 pays échantillonnés ont une espérance de vie qui se situe à plus ou moins 9.7 ans de la moyenne 73.5 ans, soit entre 63.8 ans (73.5-9.7) et 83.2 ans (73.5+9.7).**

c) Calculez le coefficient de variation. Interprétez statistiquement.

$$CV = (s/Moy) * 100 = 53,7\%$$

$$CV = (9,7/73,5) * 100 = 13,2\%$$

Le coefficient de variation relative est de 13.2%. **Autrement dit, la distribution de l'espérance de vie au sein des 10 pays est homogène puisque le coefficient de variation est inférieur à 15%.**

d) Sachant que les quartiles 1 et 3 sont respectivement de 69,8 ans et de 78 ans, calculez puis interprétez statistiquement l'intervalle interquartile.

$$IQ = Q3 - Q1.$$

$$IQ = 78 - 69,8 = 8,2.$$

L'intervalle interquartile est égal à 8.2. **Autrement dit, l'étendue de la moitié centrale de l'espérance de vie est de 8.2 ans. Encore, au moins au sein de 50% des 10 pays échantillonnés, l'espérance de vie se situe entre 69.8 et 78 ans.**

e) Si vous deviez choisir une des quatre mesures de variation tantôt calculées, laquelle choisiriez-vous comme étant la plus appropriée pour décrire la variabilité de l'espérance de vie?

J'estime que le coefficient de variation relative est plus approprié pour décrire la variabilité de l'espérance de vie au sein des 10 pays en ce sens qu'il intègre non seulement l'écart-type et la moyenne, mais aussi il offre une base de comparaison sur 100 pour mieux apprécier le degré d'homogénéité de la distribution.

5) À partir de la distribution d'âge ci-dessous d'un échantillon ($n = 70$), qui donne une moyenne de 20,03 et une médiane de 19,25,

Classes d'âge X	f	Pt milieu X	Pt X - \bar{X}	$(\text{Pt X} - \bar{X})^2$	f * $(\text{Pt X} - \bar{X})^2$
[16-18[6	17	-3,03	9,18	55,0854
[18-20[33	19	-1,03	1,06	35,0097
[20-22[20	21	0,97	0,94	18,818
[22-24[11	23	2,97	8,82	97,0299
Total (n)	70		Σ	20,00	205,943

a) Complétez le tableau ci-dessous, calculez et interprétez l'écart-type de cette distribution.

La variance d'un échantillon est égale à la somme des carrés divisée par le nombre de cas n moins 1. Donc elle est égale à $205,943/69 = 2,98$.

L'écart-type étant égal à la racine carrée de la variance ($\sqrt{2,98}$), il est de 1,73.

Environ les deux tiers des répondants (68%) sont âgés de plus ou moins 1,73 de la moyenne 20,03 ans, soit entre 18,3 et 21,8 ans.

b) Quelle hypothèse en lien avec les classes valide-t-elle l'estimation de l'écart-type?

Pour que l'estimation de l'écart-type soit valide, on fait l'hypothèse que les classes sont homogènes et surtout que les scores sont uniformément répartis dans ces classes.

c) Calculez et interprétez le coefficient d'asymétrie de cette distribution.

$$\text{Coefficient d'asymétrie} = \frac{3(\bar{X} - Md)}{s} = \frac{3(20,03 - 19,25)}{1,73} = 1,35$$

Le coefficient d'asymétrie étant de 1,35, la distribution est positivement asymétrique.

6) Le tableau suivant présente la moyenne et l'écart-type de la variable « part de financement des études par les parents » en pourcentages selon l'occupation du père. Les résultats ont été obtenus à partir d'un échantillon aléatoire constitué de 100 étudiants.

Mesures	Occupation du père			
	Chef d'entreprise	Cadres	Prof. Inter.	Employés
Part moyenne	48.22	56.94	43.89	35.12
Écart-type	41.503	40.29	40.358	38.008

a) Calculez le coefficient de variation de la variable 'part de financement des études par les parents' pour les employés. Interprétez statistiquement.

$$CV = \frac{s}{\bar{x}} * 100 = \frac{38.008}{35.12} * 100 = 108.22\%$$

Le coefficient de variation est de 108,22% : ce qui est de loin supérieur ou égal à 15%. Par conséquent, la distribution de la part de financement des études par les parents est très hétérogène chez les employés.

- b) Calculez le coefficient de variation de la variable 'part de financement des études par les parents' pour les cadres. Interprétez statistiquement.

$$CV = \frac{s}{\bar{x}} * 100 = \frac{40.293}{56.94} * 100 = 70.76\%$$

Le coefficient de variation est de 70,76%, ce qui est de loin supérieur ou égal à 15%. Par conséquent, la distribution de la part de financement des études par les parents est très hétérogène chez les cadres.

- c) Pour laquelle de ces 2 professions la variable « part de financement des études par les parents » est-elle la plus concentrée autour de la moyenne ?

La variable « part de financement des études par les parents » est plus concentrée autour de la moyenne pour les cadres que pour les employés puisque le CV pour les cadres (70.76%) est plus petit que celui des employés (108.22%).

7) Mamadou est un étudiant finissant le Bacc qui projette de s'inscrire à la maîtrise en sociologie. Pour la réalisation future de son mémoire de maîtrise, il hésite entre le choix d'une méthode qualitative et celui d'une méthode quantitative. Pour faire un choix judicieux, il décide de suivre un cours de méthode qualitative et un cours de méthode quantitative afin de choisir celle où il réussit le mieux. Les résultats sont les suivants :

- **Méthode qualitative : 71%, sachant que la moyenne et l'écart-type de la classe sont respectivement de 65 et 7**
- **Méthode quantitative : 65%, sachant que la moyenne et l'écart-type de la classe sont respectivement de 60 et 5**

- a) Pour faire un choix éclairé, pourquoi Mamadou ne peut se limiter à comparer les notes obtenues en méthode qualitative et en méthode quantitative?

Il ne peut se fier aux notes obtenues, car les deux matières sont différentes par essence et ne sont pas notées de la même façon.

- b) Quelle est la bonne décision qu'il doit prendre en référence à son domaine d'excellence ? Justifiez la décision.

Pour prendre la bonne décision en référence à son domaine d'excellence, il doit standardiser ses notes de façon à mettre à plat les différences artificielles liées aux matières et au système de notation.

$$Z = (X_i - \text{Moy}) / s$$

Pour la note obtenue en méthode qualitative, $Z = (71 - 65) / 7 = 0,86$

Pour la note obtenue en méthode quantitative, $Z = (65 - 60) / 5 = 1$

Pour la réalisation future de son mémoire de maîtrise, je conseille à Mamadou de choisir la méthode quantitative, car sa note se situe à 1 écart-type au dessus de la

moyenne de sa classe.

8) La moyenne d'un examen est de 60% et l'écart type de 10%. La distribution des notes à cet examen s'apparente à une distribution normale

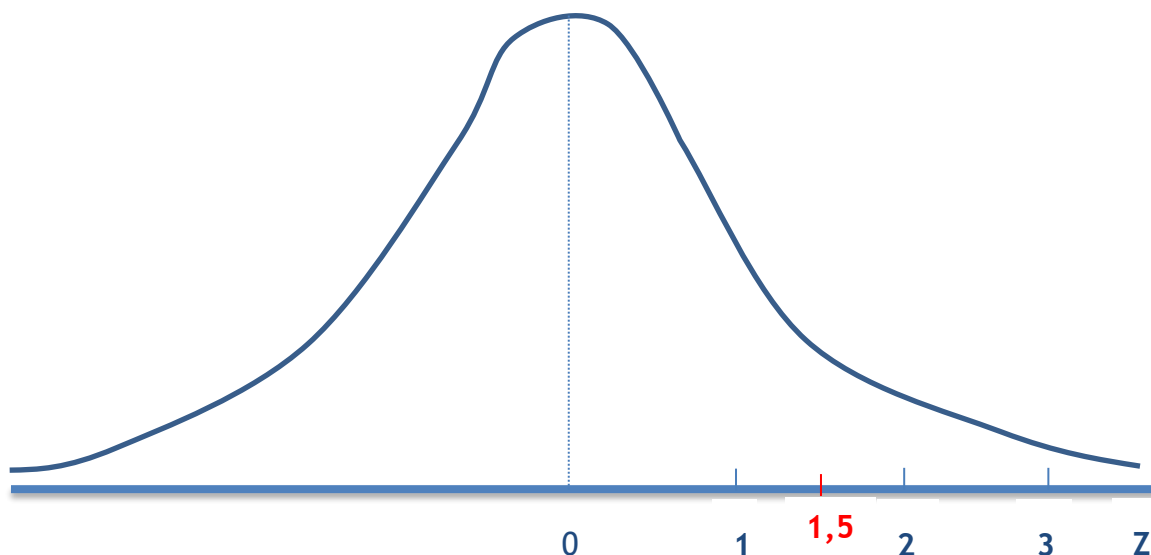
- a) Le score-z de la note de Khady est de 1,5. Pourquoi les conditions d'utilisation de la loi normale sont-elles réunies pour représenter la note de Khady ?

D'une part, les notes suivent une distribution normale : autrement dit, elles sont distribuées normalement autour de la moyenne 60%.

D'autre part, la note de Khady étant transformée en score z (1,5), elle peut être approchée et représentée à l'aide de la loi normale.

NB : Pour approcher un phénomène à l'aide de la loi normale, au moins deux conditions doivent être satisfaites : la variable continue aléatoire est distribuée normalement et les scores-X sont standardisés. La standardisation des scores-X est obtenue en les transformant en scores-Z (centrage et réduction) ☺

- b) Représentez la cote-z de Khady sur cette courbe normale et donnez son sens.



Le score-z de la note de Khady est de 1,5: **ce qui signifie que sa note se situe à 1,5 écart-type au-dessus de la moyenne de sa classe.**

- c) En vous référant à la courbe normale, combien de points Khady a-t-elle de plus que la moyenne? Quelle est donc sa note réelle?

Si

$$Z = \frac{X - \bar{X}}{S}$$

Alors

$$ZS = X - \bar{X} \quad X = ZS + \bar{X}$$

$X = 60 + (1,5 * 10) = 75$. La note de Khady est de 75%, soit 15% de plus que la moyenne de sa classe 60%.

9) En vous référant à la loi normale, calculez la surface (ou proportion) comprise

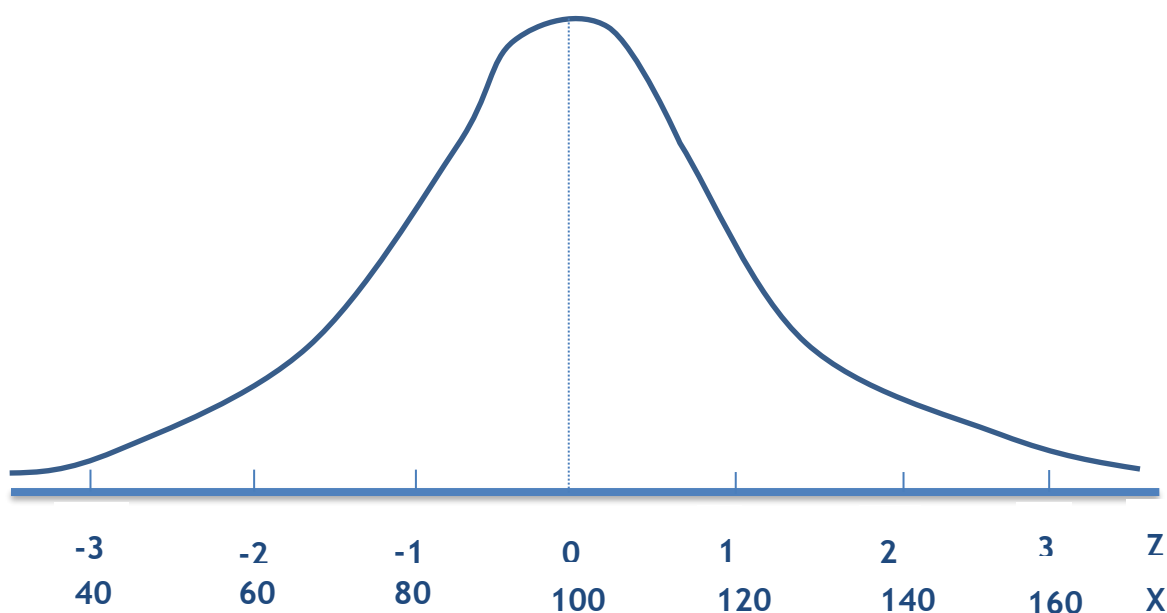
sous la courbe normale pour les intervalles de scores standardisés z_i suivants

	En d'autres termes	Proportion	Calcul de la proportion si nécessaire
a)	de $-\infty$ à 2,4	$P(z < 2,4)$	0,9918
b)	de $-\infty$ à -2,0	$P(z < -2)$	0,0228
c)	de -1 à 1	$P(-1 < z < 1)$	0,8413-0,1587
d)	de 0 à 2,2	$P(0 < z < 2,2)$	0,9861-0,5
e)	de 2,2 à $+\infty$	$P(z > 2,2)$	0,0139

10) Un test d'intelligence dans une université produit les paramètres suivants:

$$\begin{aligned}\mu &= 100 \\ \sigma &= 20 \\ X &\sim N(100; 20)\end{aligned}$$

10.1. Décrivez la distribution des scores- X et des scores- Z en les représentant sur la courbe normale à l'aide des paramètres ci-dessus.



10.2. En vous servant de la loi normale, calculez la proportion (en %) d'étudiants qui obtient un quotient intellectuel (QI) correspondant aux caractéristiques suivantes :

a) Inférieur à 80

Trouvons d'abord le score z correspondant au score- X 80 (transformation)

$$Z = (80-100) / 20 = -1$$

Puis, cherchons $P(z < -1)$ ou bien $P(z > 1)$: soit 0,1587

La proportion ou pourcentage d'étudiants ayant un QI inférieur à 80 est de 16%.

b) Inférieur à 90

Trouvons d'abord le score z correspondant au score-X 90

$$Z = (90-100) / 20 = -0,5$$

Puis, cherchons $P(z < -0,5)$ ou bien $P(z > 0,5)$: soit 0,3085

La proportion ou pourcentage d'étudiants ayant un QI inférieur à 90 est de 31%.

c) Inférieur à 100

Trouvons d'abord le score z correspondant au score X 100

$$Z = (100-100) / 20 = 0$$

Puis, cherchons $P(z < 0)$: soit 0,5

La proportion ou pourcentage d'étudiants ayant un QI inférieur à 100 est de 50%.

d) Supérieur à 120

Trouvons d'abord le score z correspondant au score X 120

$$Z = (120-100) / 20 = 1$$

Puis, cherchons $P(z > 1)$: soit 0,1587

La proportion d'étudiants ayant un QI supérieur à 120 est de 16%.

e) Supérieur à 130

Trouvons d'abord le score z correspondant au score X 130

$$Z = (130-100) / 20 = 1,5$$

Puis, cherchons $P(z > 1,5)$: soit 0,0668

Le pourcentage d'étudiants ayant un QI supérieur à 130 est de 7%.

f) Compris entre 80 et 120

$$P(80 < X < 120) = P(-1 < z < 1)$$

$$= P(z < 1) - P(z < -1) = 0,8413 - 0,1587 = 0,6826$$

Le pourcentage d'étudiants ayant un QI compris entre 80 et 120 est de 68%.

g) Compris entre 90 et 140

$$P(90 < X < 140) = P(-0,5 < z < 2)$$

$$= P(z < 2) - P(z < -0,5) = 0,9772 - 0,3085 = 0,6687$$

La proportion d'étudiants ayant un QI compris entre 90 et 140 est de 67%.

h) Compris entre 100 et 140

$$P(100 < X < 140) = P(0 < z < 2)$$

$$= P(z < 2) - P(z < 0) = 0,9772 - 0,5 = 0,4772$$

La proportion ou pourcentage d'étudiants ayant un QI compris entre 100 et 140 est de 48%.