Standardisation - Loi normale

J-M., GALHARRET

LMJL, Faculté de Psychologie, Nantes Université

Standardisation des données
Scores z

Position du problème

- Un score brut n'indique rien. Si on passe un test pour mesurer notre créativité et qu'on obtient un score de $X = 30$... on ne peut rien conclure.
- Par contre si on sait que la moyenne des scores de créativité est $\mu = 15$, alors on peut dire au moins que notre score est supérieur au score "typique". Mais ce n'est pas encore suffisant ....

Allons plus loin

- Si de plus on sait que l'écart type des scores est 11 alors notre score est compris entre $\mu + \sigma$ et $\mu + 2 * \sigma$ et on peut même chercher précisement à quelle déviation (en écart type) on se trouve de la moyenne $X = \mu + Z\sigma$ ce qui donne

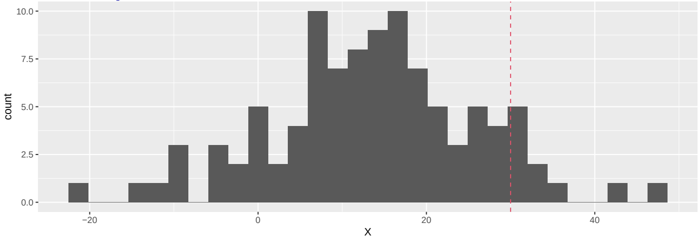
$$Z = \frac{X - \mu}{\sigma}.$$

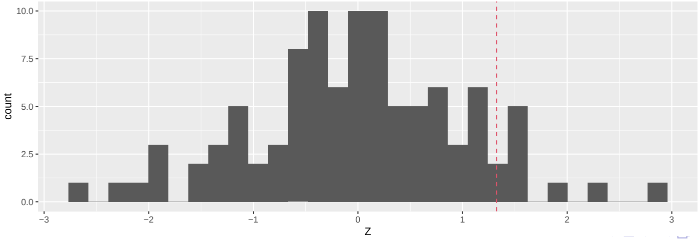
Dans notre exemple $Z = 1.4$ donc d'après BT et si on considère que la répartion des scores autour de la moyenne est symétrique on fait parti des 24.5 % des scores les plus créatifs.

Définition des scores z

- Etant donnée une variable X_i de moyenne μ et d'écart type σ on appelle score standardisé ou normalisé ou score z le score Z_i défini par :
$$Z_i = \frac{X_i - \mu}{\sigma}.$$
- Si on ne connait pas (μ, σ) et qu'on possède un échantillon de valeurs $(X_i)_{i=1,\ldots,n}$ alors le score Z_i est défini par
$$Z_i = \frac{X_i - \bar{X}}{s},$$
où s est l'écart type corrigé.

Pour aller plus loin





Autre exemple

- On a également été testé en terme d'intelligence émotionnelle et on a obtenu un score de 175 avec un questionnaire étalonné de moyenne $\mu = 150$ et d'écart type $\sigma = 20$.
- A-t-on un meilleur score en termes d'IE ou de créativité ?

Solution : On calcule le score $Z_{IE} = 1.2$ et on avait $Z_{Crea} = 1.4$. $\rightsquigarrow$ on est plus proche de la moyenne en terme d'intelligence émotionnelle que de créativité. On est donc plutôt créatif !

Résumé

- Les scores standardisés permettent de comparer le score d'un individu par rapport aux caractéristiques d'une population ou d'un échantillon (moyenne et écart type).
- les scores standardisés permettent de comparer des scores entre eux même si ils ne sont pas sur la même échelle.
- Quelle que soit la variable de départ X, la variable standardisée Z a pour moyenne 0 et pour écart type 1.

Loi normale centrée réduite

Loi de probabilité

On peut associer une loi de probabilité à une variable numérique .

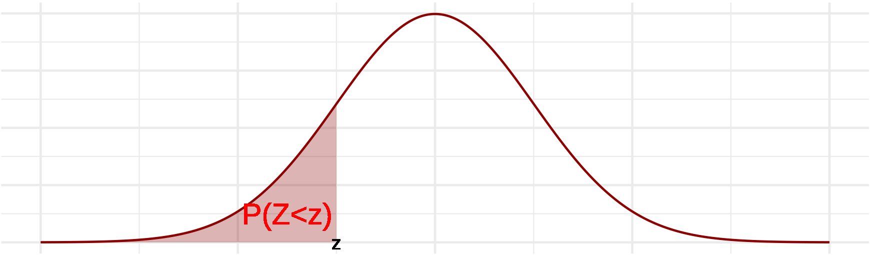
Exemple On jette deux fois une pièce et on définit X comme le nombre de fois où la pièce est tombée sur pile. X peut prendre les valeurs 0,1,2 et on a

- $\mathbb{P}(X = 0) = \frac{1}{4}$
- $\mathbb{P}(X = 1) = \frac{1}{2}$
- $\mathbb{P}(X = 2) = \frac{1}{4}$

- Lorsque la variable est continue c'est plus compliqué : il faut définir pour toute valeur de x de X on va définir la valeur $\mathbb{P}(X < x)$.
- L'une des lois les plus utilisées est la loi normale centrée réduite...
- Si une variable Z suit une loi normale centrée réduite on note $Z \sim \mathcal{N}(0, 1)$

Représentation graphique

La courbe de $Z \sim \mathcal{N}(0,1)$ (courbe gaussienne) est :



Propriétés :

La loi normale centrée réduite :

- est symétrique, donc sa médiane est égale à sa moyenne (=0).
- a une très faible probabilté de prendre des petites (ou des grandes) valeurs !
- est telle que pour tout z on a $\mathbb{P}(Z < z) = \mathbb{P}(Z \leq z)$, autrement dit $\mathbb{P}(Z = z) = 0$

Table

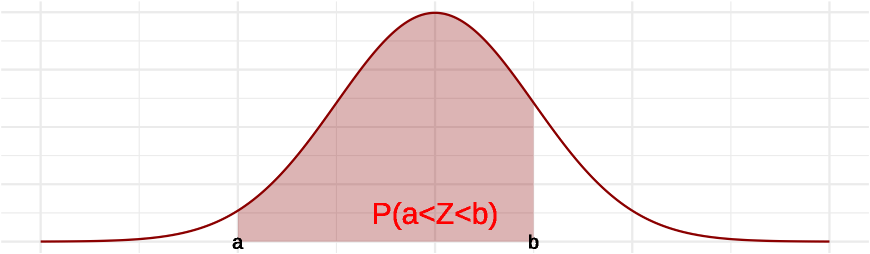
La table ci-dessous donne les valeurs de $\mathbb{P}(Z \leq z)$ pour $Z \sim \mathcal{N}(0, 1)$, z étant égal à la somme de la valeur de la ligne et celle de la colonne.

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.500	0.504	0.508	0.512	0.516	0.520	0.524	0.528	0.532	0.536
0.5	0.691	0.695	0.698	0.702	0.705	0.709	0.712	0.716	0.719	0.722
1	0.841	0.844	0.846	0.848	0.851	0.853	0.855	0.858	0.860	0.862
1.5	0.933	0.934	0.936	0.937	0.938	0.939	0.941	0.942	0.943	0.944
2	0.977	0.978	0.978	0.979	0.979	0.980	0.980	0.981	0.981	0.982

$$\mathbb{P}(Z > z) = 1 - \mathbb{P}(Z \leq z)$$

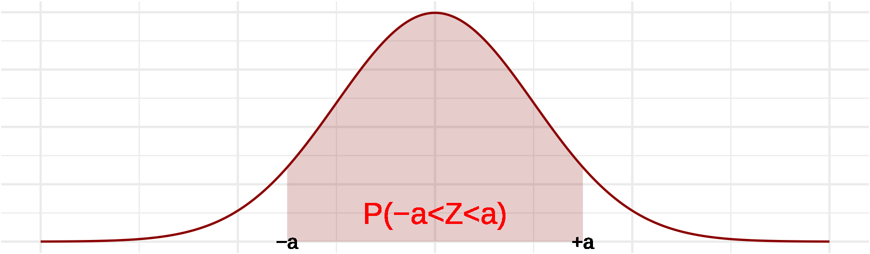
Exemple Calculer $\mathbb{P}(Z > 0.75)$

$$\mathbb{P}(a < Z < b) = \mathbb{P}(Z \leq b) - \mathbb{P}(Z \leq a)$$



Exemple Calculer $\mathbb{P}(1.2 < Z < 1.35)$

$$\mathbb{P}(-a < Z < a) = 2 \times \mathbb{P}(Z < a) - 1$$

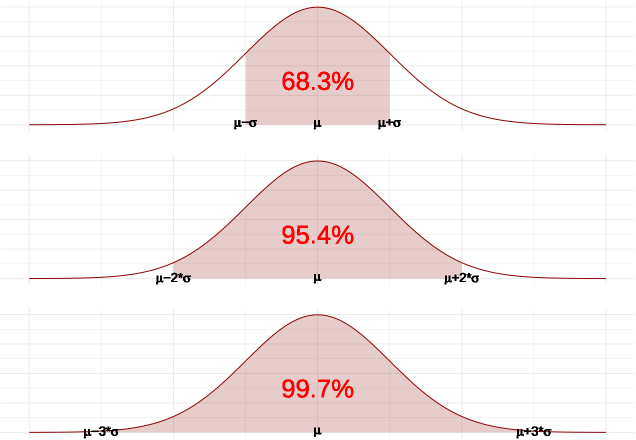


Exemple Calculer $\mathbb{P}(-1.2 < Z < 1.2)$

Lois normales $\mathcal{N}(\mu, \sigma)$:

- Une loi normale $X \sim \mathcal{N}(\mu, \sigma)$ est une variable dont le score $Z \sim \mathcal{N}(0, 1)$.
- Une loi normale $X \sim \mathcal{N}(\mu, \sigma)$ a pour moyenne μ et pour écart type σ. Donc une loi normale est symétrique autour de sa moyenne.

Intervalles à connaitre



Retour sur la créativité :

Si on suppose que le score de créativité est $X \sim \mathcal{N}(15, 11)$ alors le score $Z = 1.4$ correspondant à $X = 30$ vérifie :

$\mathbb{P}(Z > 1.4) = 8.1$ %.