



Compte-rendu d'analyse

valant évaluation dans le cadre de :

Diplôme : Master statistique pour l'évaluation et prévision, 1^{ère} année

Année universitaire : 2022-2023

Module d'enseignement : SEP0831

Responsable : Philippe Regnault

Comptant pour : 50 %

La popularité des musiques sur la plateforme Spotify

Elif ERTAS

`elif.ertas@etudiant.univ-reims.fr`

Métriques

Finalisé le : 25 avril 2023

Page(s) : 13

Références(s) : 0

Figure(s) : 4

Table(s) : 0

Théorème(s) : 0

Résumé : À partir d'une base de donnée provenant directement de la plateforme Spotify, on veut expliquer la popularité des musiques. On étudiera donc le phénomène à travers différentes variables tel que le type de groupe, la danceabilité et l'acoustique.

Mots-clés : Spotify, Top music, Popularity.

Matériel supplémentaire :

git: <https://github.com/pregnault/urcadown>

📁: <http://organizyourmusic.playlistmachinery.com>

Table des matières

1	Description des données	3
1.1	La plateforme de streaming Spotify	3
1.2	L'objectif de l'étude	3
1.3	Les données utilisées	3
2	Les différents types d'artistes	4
2.1	La répartition des types d'artistes	4
2.2	La popularité entre les différents types d'artistes	4
3	La danceabilité des musiques	5
3.1	La popularité des musiques en fonction de leur danceabilité	6
4	L'acoustique des musiques	6
4.1	La popularité des musiques en fonction de leur acoustique	6
5	Pour aller plus loin	7
6	Conclusion	7
	Remerciements	7
A	Annexes	8
A.1	Structure de la base de donnée	8
A.2	Tests de normalité entre le type d'artiste et la popularité	8
A.3	Test de l'égalité des variances entre le type d'artiste et la popularité	10
A.4	Test de comparaison des médianes entre le type d'artiste et la popularité	10
A.5	Comparaison deux à deux des groupes entre le type d'artiste et la popularité	11
A.6	Test de normalité entre la danceabilité et la popularité	11
A.7	Tests de non-corrélation monotone entre la danceabilité et la popularité	11
A.8	Test de normalité entre l'acoustique et la popularité	12
A.9	Tests de non-corrélation monotone entre l'acoustique et la popularité	12
	Bibliographie	13

1. Description des données

1.1. La plateforme de streaming Spotify

Depuis des décennies, la musique se trouve partout dans notre quotidien, encore plus aujourd'hui avec toutes les nouvelles technologies et en particulier les réseaux sociaux. Il existe une multitude de plateformes de streaming musical dans lesquelles on peut découvrir des musiques du monde entier. Cependant, il y a toujours des musiques plus populaires que d'autres, d'ailleurs les classements peuvent être très différents d'un pays à l'autre. La culture musicale peut varier d'une zone à l'autre, c'est donc pour cela que notre étude se fera uniquement sur le classement mondial des musiques. Spotify est l'un des services de streaming musical les plus connues, cette plateforme permet une écoute instantanée de diverses musiques. De plus, c'est une plateforme gratuite, ce qui permet d'avoir encore plus d'utilisateurs. La plateforme compte aujourd'hui plus de 500 millions d'utilisateurs mensuels, répartis dans le monde entier et de tout âge.

1.2. L'objectif de l'étude

La présente étude vise à mettre en évidence l'impact de divers facteurs sur la popularité d'une musique dans le top 100 annuel mondial sur la plateforme de streaming Spotify. Précisément, on cherche à identifier des différences significatives de la popularité entre le type de groupe, la danceabilité ou l'acoustique d'une musique.

1.3. Les données utilisées

Cette étude est possible grâce à une base de données¹ regroupant les 100 meilleures musiques de chaque année entre 2010 et 2019 ce qui nous fait un total de 1000 musiques. Cette base de données provient d'un site appartenant à Spotify² qui permet de récolter les informations de n'importe quelles musiques sur leur plateforme.

Précisément, parmi les informations (variables) recueillies, citons³ :

- **title** : le titre de la musique (on y trouve également le nom des collaborateurs s'il y en a) ;
- **artist** : Le nom de l'artiste ou du groupe principal qui interprète la musique ;
- **top_genre** : Le genre dominant de la musique (ici 132 genres différents) ;
- **bpm** : Le nombre de battements par minute (=tempo) ;
- **nrngy** : L'énergie de la musique (de 0 à 100, plus elle est élevée et plus la musique est énergétique) ;
- **dnce** : La danceabilité de la musique (de 0 à 100, plus elle est élevée et plus il est facile de danser) ;
- **val** : La valence de la musique (de 0 à 100, plus elle est élevée et plus la musique est joyeuse) ;
- **acous** : L'acoustique de la musique (de 0 à 100) ;
- **pop** : La popularité de la musique sur la plateforme Spotify (de 0 à 100) ;
- **artist_type** : Le type d'artiste, c'est-à-dire en Solo, Duo, Trio ou Groupe ;

1. Il s'agit de données officielles, issues directement de la plateforme Spotify.

2. <http://organizyourmusic.playlistmachinery.com/> la base de donnée est constituée automatiquement en précisant pour quelle(s) playlist(s) de la plateforme Spotify vous voulez extraire les informations.

3. Il y a d'autres variables présentes dans la base de données qui ne sont pas citées car elles ne nous intéressent pas, voir l'annexe 1 pour plus d'informations.

2. Les différents types d'artistes

2.1. La répartition des types d'artistes

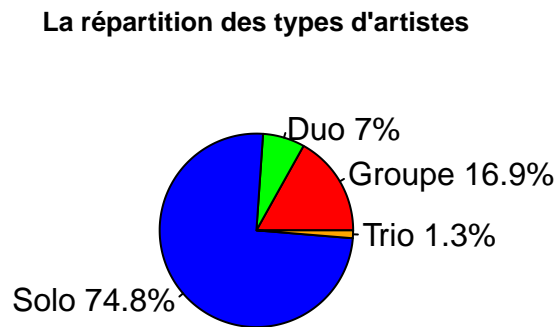


FIGURE 1 – Répartition des types d'artistes.

Cohorte : L'ensemble des 1000 musiques de la base de données.

Lecture : On observe que le type d'artiste qui domine est le Solo à 74.8% suivi par le type Groupe à 16.9%, le type Duo à 7% puis pour finir les trio à 1.3%

Le type d'artiste qui domine est le soliste, cette dominance est dû en particulier aux nombreuses collaborations entre artistes. En effet le type "Solo" prend en compte les titres de musiques faites par un seul interprète tel que Rihanna mais aussi les titres de musiques faites en collaboration où l'artiste principal est un Solo tel que "Sucker for Pain" qui est un titre de Lil Wayne (Solo) en collaboration avec Wiz Khalifa (Solo), Imagine Dragons (Groupe de 4) et d'autres artistes. En général, ce sont les artistes solo qui font le plus de collaborations.

2.2. La popularité entre les différents types d'artistes

En février 2021, le site **Europavox** sort publiait un article ⁴ [Fer21] faisant écho que les artistes solo auraient tendance à produire des chansons plus populaires que les groupes, notamment grâce aux nombreuses collaborations qui ont énormément du succès. Cette popularité serait dû aux mélanges des genres musicales et donc une manière d'atteindre un nouveau public pour tous les collaborateurs du titre de musique.

On observe pas de grosses différences entre chaque type d'artiste comme décrit ⁵ par la figure 2.

4. <https://www.europavox.com/fr/news/how-the-featured-artist-is-excelling-across-europe/>

5. Il y a 13 titres qui sortent de la normal pour le type Solo, ce qui signifie qu'il y a 13 titres de musique d'artistes solo qui ont une popularité nettement plus faible que les autres. On a une moyenne assez similaire les uns aux autres. De plus on observe un défaut de symétrie pour le type d'artiste Trio.

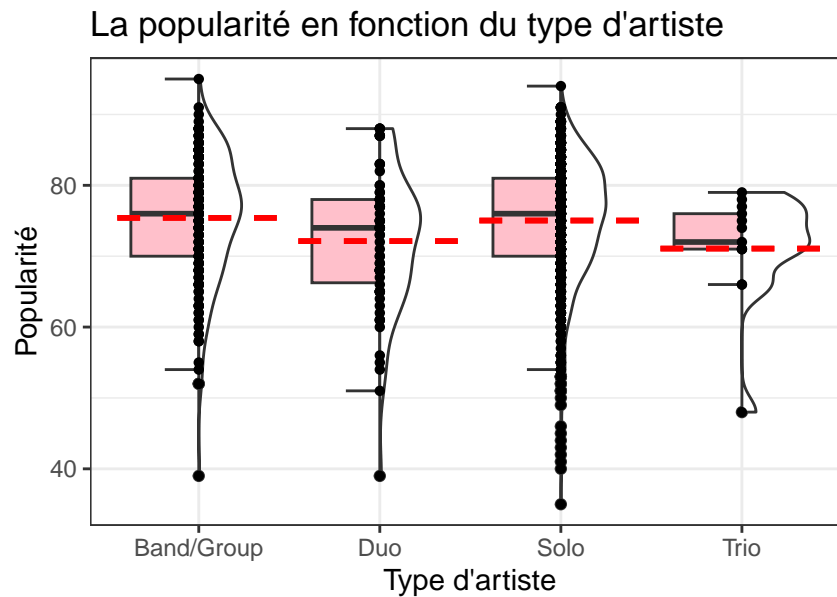


FIGURE 2 – Popularité des musiques en fonction de leurs types d'artiste.

Cohorte : L'ensemble des 1000 musiques de la base de données.

Lecture : On observe une moyenne de 75 de popularité pour les Groupes.

Après analyses et résultats des tests⁶, on a une tendance de popularité élevée dans les Groupes et Solo par rapport à Duo mais on ne peut se prononcer entre Groupe et Solo au vu de nos données. On suppose que la non-significativité⁷ avec le groupe Trio est dû au manque de données (1.3% pour Trio)

On conclut qu'il y a bien une différence de popularité entre les types d'artistes, cependant on ne peut pas dire si un type est significativement plus populaire que tous les autres types.

3. La danceabilité des musiques

La danceabilité est mesurée à l'aide d'algorithmes d'analyse de la musique. Les plateformes de streaming de musique, comme Spotify, utilisent des algorithmes d'apprentissage automatique pour calculer la danceabilité des morceaux de musique en fonction de diverses caractéristiques, telles que le tempo, le rythme, la densité de la batterie et la stabilité rythmique. Les algorithmes de calcul de la danceabilité sont entraînés sur des données d'analyse musicale et peuvent être affinés en fonction des préférences des auditeurs, en utilisant des données d'écoute et des commentaires sur la musique.

La danceabilité est une mesure importante pour les DJ et les producteurs de musique, car elle leur permet de sélectionner des morceaux qui feront danser le public. Elle est également utilisée par les analystes de données pour comprendre les préférences des auditeurs et pour recommander des morceaux similaires en fonction de la danceabilité.

On peut alors se demander si la popularité des musiques n'est pas dû à leur danceabilité. En effet, dans notre quotidien, on entend plus souvent des musiques qui donnent envie de danser, que ce soit à la radio, à la télévision ou sur les réseaux sociaux. Cependant, ce n'est qu'un environnement propre à soi même, donc il n'y a pas d'étude ou d'article visant à imposer que la popularité d'une musique est impactée majoritairement par sa danceabilité. Elle peut tout de même avoir un rôle important, c'est ce que nous allons étudier.

6. Des tests de normalité ont été réalisés dans la partie A.2 ; un test d'égalité des variances dans la partie A.3 ; un test de comparaison des médianes dans la partie A.4 ; une comparaison des groupes deux à deux dans la partie A.5 ;

7. Le niveau de significativité de tous les tests effectués a été fixé à 5%.

3.1. La popularité des musiques en fonction de leur danceabilité

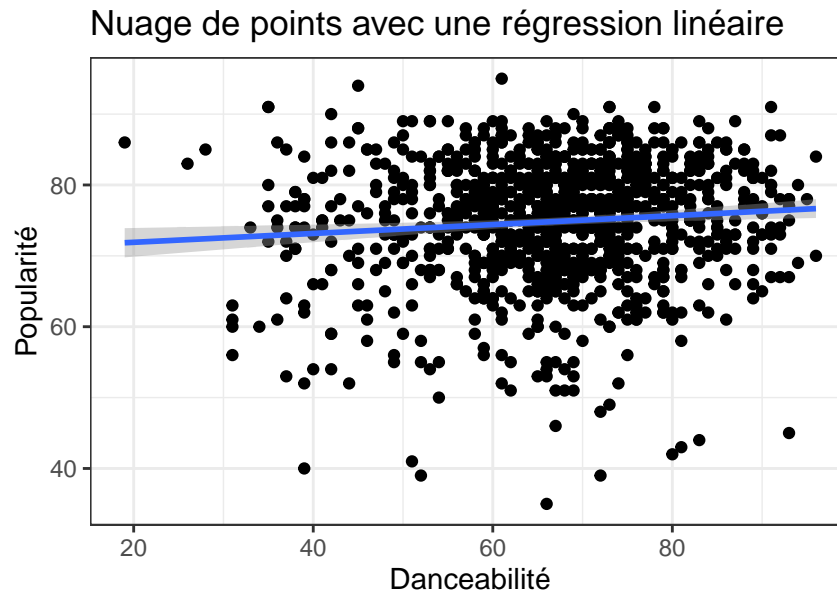


FIGURE 3 – Popularité des musiques en fonction de leur danceabilité. **Cohorte** : L'ensemble des 1000 musiques de la base de données. **Lecture** : Un titre de musique avec 80 de danceabilité a environ 75 de popularité.

Par la figure 3 on n'observe aucune relation particulière entre la popularité et la danceability. Après plusieurs tests⁸ on conclut qu'il y a existence d'une relation monotone de dépendance entre la popularité et la danceabilité

4. L'acoustique des musiques

L'acoustique d'une musique se réfère aux caractéristiques sonores de la musique, telles que le timbre, le tempo, la hauteur, le rythme, le loudness, la durée et l'énergie sonore. Ces caractéristiques sonores peuvent être utilisées pour identifier le genre musical, le style de production ou l'instrumentation de la musique. Par exemple, une musique acoustique est généralement jouée avec des instruments acoustiques, tels que la guitare acoustique, la contrebasse, le piano acoustique ou la batterie acoustique. En revanche, une musique électronique est souvent produite à l'aide de synthétiseurs, de boîtes à rythmes et d'autres équipements électroniques. L'acoustique d'une musique peut être mesurée à l'aide d'algorithmes d'analyse acoustique, qui peuvent quantifier différentes caractéristiques sonores de la musique.

4.1. La popularité des musiques en fonction de leur acoustique

Par la figure 4 on n'observe aucune relation particulière entre la popularité et l'acoustique. Après plusieurs tests⁹ on conclut qu'il y a existence d'une relation monotone de dépendance entre la popularité et la danceabilité

8. Un test de normalité a été réalisé dans la partie A.6 ; des tests de non-corrélation monotone sont réalisés dans la partie A.7 ;

9. Un test de normalité a été réalisé dans la partie A.8 ; des tests de non-corrélation monotone sont réalisés dans la partie A.9 ;

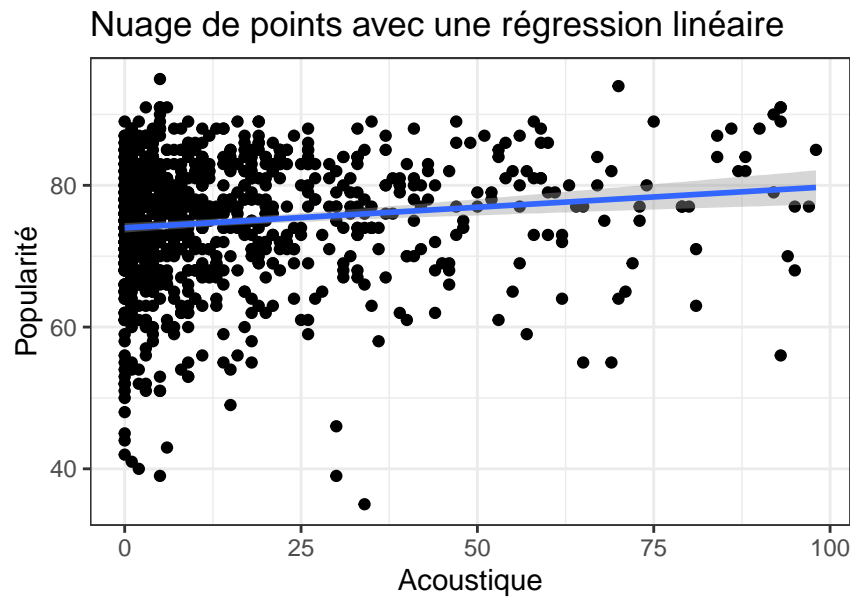


FIGURE 4 – Popularité des musiques en fonction de leur acoustique. **Cohorte** : L'ensemble des 1000 musiques de la base de données. **Lecture** : Un titre de musique avec 80 d'acoustique a environ 80 de popularité.

5. Pour aller plus loin

Il serait intéressant de collecter ces données pour un seul artiste, en ayant quand même plusieurs dizaines de titres pour avoir des résultats concluants. On pourrait alors s'intéresser à cet artiste et comprendre pourquoi certains de ses titres sont plus ou moins populaires que d'autres.

On pourrait également collecter ces données pour un genre de musique en particulier comme le hip-hop par exemple, on aurait des titres de musiques du même environnement. On pourrait les prendre avec des dates de sortie rapprochées pour observer s'il y a un système de concurrence mis en avant et comprendre pourquoi tel titre est plus populaire qu'un autre.

6. Conclusion

Pour conclure, nous avons mis en évidence que la popularité d'une musique ne peut être expliquée dans sa globalité pour cette cohorte de titres de musiques. Des musiques avec des caractéristiques complètement différentes peuvent se trouver au même seuil de popularité. Une musique n'est donc pas vraiment populaire par son nombre d'interprète, son acoustique ou sa danceabilité. On peut alors toujours se demander comment une musique peut être populaire ? Il y a un effet de mode et le partage exerce également une forte influence sur la popularité d'une musique. On a par exemple le réseau social TikTok qui a permis à des milliers de titres de musiques de revenir à la mode ou même de nouvelles musiques à devenir connues.

Remerciements

Merci à M. Regnault pour son enseignement et ses conseils.

A. Annexes

A.1. Structure de la base de donnée

```
'data.frame': 1000 obs. of 17 variables:
 $ title      : chr  "STARSTRUKK (feat. Katy Perry)" "My First Kiss (feat. Ke$ha)" "I Need A
 $ artist     : chr  "3OH!3" "3OH!3" "Aloe Blacc" "B.o.B" ...
 $ top_genre  : chr  "dance pop" "dance pop" "pop soul" "atl hip hop" ...
 $ year_released: num  2009 2010 2010 2010 2010 ...
 $ added      : chr  "2022-02-17" "2022-02-17" "2022-02-17" "2022-02-17" ...
 $ bpm        : num  140 138 95 93 104 82 128 92 146 109 ...
 $ nrgy       : num  81 89 48 87 85 93 81 52 59 84 ...
 $ dnce       : num  61 68 84 66 69 55 82 60 50 64 ...
 $ dB         : num  -6 -4 -7 -4 -6 -4 -8 -7 -5 -5 ...
 $ live       : num  23 36 9 4 9 35 60 31 11 6 ...
 $ val        : num  23 83 96 38 74 79 44 41 8 42 ...
 $ dur        : num  203 192 243 180 268 196 308 258 218 221 ...
 $ acous      : num  0 1 20 11 39 1 7 18 51 1 ...
 $ spch       : num  6 8 3 12 5 34 7 37 3 4 ...
 $ pop        : num  70 68 72 80 79 71 75 71 87 86 ...
 $ top_year   : num  2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...
 $ artist_type: chr  "Duo" "Duo" "Solo" "Solo" ...
```

A.2. Tests de normalité entre le type d'artiste et la popularité

```
$'Band/Group'
```

```
Shapiro-Wilk normality test
```

```
data: X[[i]]
```

```
W = 0.96646, p-value = 0.000419
```

```
$Duo
```

```
Shapiro-Wilk normality test
```

```
data: X[[i]]
```

```
W = 0.96296, p-value = 0.03656
```

```
$Solo
```

```
Shapiro-Wilk normality test
```

```
data: X[[i]]
```

```
W = 0.94814, p-value = 1.552e-15
```


\$Trio

Shapiro-Wilk normality test

data: X[[i]]

W = 0.78376, p-value = 0.0044

\$'Band/Group'

Jarque-Bera Normality Test

data: X[[i]]

JB = 21.918, p-value = 1.74e-05

alternative hypothesis: greater

\$Duo

Jarque-Bera Normality Test

data: X[[i]]

JB = 7.3038, p-value = 0.02594

alternative hypothesis: greater

\$Solo

Jarque-Bera Normality Test

data: X[[i]]

JB = 216.18, p-value < 2.2e-16

alternative hypothesis: greater

\$Trio

Jarque-Bera Normality Test

data: X[[i]]

JB = 13.465, p-value = 0.001191

alternative hypothesis: greater

\$'Band/Group'

Anderson-Darling normality test

data: X[[i]]

A = 1.1928, p-value = 0.004006

\$Duo

Anderson-Darling normality test

```
data:  X[[i]]  
A = 0.52667, p-value = 0.1733
```

\$Solo

Anderson-Darling normality test

```
data:  X[[i]]  
A = 7.6109, p-value < 2.2e-16
```

\$Trio

Anderson-Darling normality test

```
data:  X[[i]]  
A = 0.94665, p-value = 0.01139
```

D'après le test de Anderson-Darling le type d'artiste Duo a une p-value de $0.1733 > 0.05$ donc on accepterait l'hypothèse de normalité pour ce type. Or il suffit qu'un test pour un seul type rejette l'hypothèse pour faire la procédure non paramétrique : Tout les groupes ont une p-value significative dans au moins un des tests, ici les types Group, Trio et Solo rejette l'hypothèse de normalité dans tout les tests effectués. Le type Duo rejette l'hypothèse de normalité d'après les tests de Shapiro-Wilk et Jarque-Berra On rejette donc l'hypothèse de normalité pour tous les types.

A.3. Test de l'égalité des variances entre le type d'artiste et la popularité

Fligner-Killeen test of homogeneity of variances

```
data:  pop by artist_type  
Fligner-Killeen:med chi-squared = 3.6607, df = 3, p-value = 0.3005
```

On a une p-value = $0.3005 > 0.05$ non significative donc on accepte l'égalité des variances

A.4. Test de comparaison des médianes entre le type d'artiste et la popularité

Kruskal-Wallis rank sum test

```
data:  pop by artist_type  
Kruskal-Wallis chi-squared = 11.319, df = 3, p-value = 0.01012
```

On a une p-value = 0.01012 significative donc on va devoir comparer deux-à-deux les groupes.

A.5. Comparaison deux à deux des groupes entre le type d'artiste et la popularité

```
# A tibble: 6 x 6
  group1    group2      p p.adj p.format p.signif
  <chr>    <chr>    <dbl> <dbl> <chr>    <chr>
1 Duo      Solo      0.00598 0.036 0.0060  **
2 Duo      Band/Group 0.00970 0.058 0.0097  **
3 Duo      Trio      0.807   1     0.8067  ns
4 Solo     Band/Group 0.631   1     0.6311  ns
5 Solo     Trio      0.0597  0.36  0.0597  ns
6 Band/Group Trio      0.0646  0.39  0.0646  ns
```

On observe une significativité entre Duo - Solo et Duo - Group. On en conclut que la popularité pour le type Solo est significativement plus grande que pour le type Duo et que la popularité pour le type Groupe est significativement plus grande pour que pour le type Duo

A.6. Test de normalité entre la danceabilité et la popularité

Shapiro-Wilk normality test

```
data: Z
W = 0.95874, p-value = 3.501e-16
```

p-value = $3.50e-16 < 0.05$ significative donc on rejette l'hypothèse de normalité

A.7. Tests de non-corrélation monotone entre la danceabilité et la popularité

Kendall's rank correlation tau

```
data: Spotify$pop and Spotify$dnce
z = 2.7651, p-value = 0.005691
alternative hypothesis: true tau is not equal to 0
sample estimates:
      tau
0.06005483
```

```
Warning in cor.test.default(Spotify$pop, Spotify$dnce, method = "spearman"):
Impossible de calculer la p-value exacte avec des ex-aequos
```

Spearman's rank correlation rho

```
data: Spotify$pop and Spotify$dnce
S = 152104509, p-value = 0.005696
alternative hypothesis: true rho is not equal to 0
sample estimates:
```

```
rho
0.08737204
```

p-value singificative pour les deux tests

A.8. Test de normalité entre l'acoustique et la popularité

```
Shapiro-Wilk normality test
```

```
data: Z
W = 0.8955, p-value < 2.2e-16
```

p-value < 2.2e-16 < 0.05 significative donc on rejette l'hypothèse de normalité

A.9. Tests de non-corrélation monotone entre l'acoustique et la popularité

```
Kendall's rank correlation tau
```

```
data: Spotify$pop and Spotify$acous
z = 4.8058, p-value = 1.542e-06
alternative hypothesis: true tau is not equal to 0
sample estimates:
tau
0.1057238
```

```
Warning in cor.test.default(Spotify$pop, Spotify$acous, method = "spearman"):
Impossible de calculer la p-value exacte avec des ex-aequos
```

```
Spearman's rank correlation rho
```

```
data: Spotify$pop and Spotify$acous
S = 141546595, p-value = 1.688e-06
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.1507196
```

p-value singificative pour les deux tests

Webographie

- [Fer21] FERREIRA AMORIM ABIASSI. *Collaborations musicales : la puissance du featuring* - *Europavox*. Collaborations musicales : la puissance du featuring - Europavox. 11 fév. 2021. URL : <https://www.europavox.com/fr/news/how-the-featured-artist-is-excelling-across-europe/> (visit  le 25/04/2023).