

Facteurs influençant la durée d'écoute musicale quotidienne des étudiants

1. Introduction et problématique

L'étude explore les déterminants de la durée quotidienne d'écoute musicale chez les étudiants français, population particulièrement exposée aux pratiques numériques audiovisuelles. Les recherches existantes montrent que 55% des jeunes écoutent plus de 2 heures quotidiennes de musique, avec des variations liées au contexte environnemental et aux habitudes d'étude.

La problématique centrale s'articule autour de l'interaction entre les variables comportementales, technologiques et psychologiques :

- ♪ Comment le profil démographique (âge, sexe) module-t-il les pratiques d'écoute ?
- ♪ Quel impact ont les usages fonctionnels de la musique (concentration vs détente) sur sa durée d'utilisation ?
- ♪ Existe-t-il une corrélation entre le support technologique privilégié et l'engagement temporel ?

Cette recherche s'inscrit dans un contexte où, d'après des études, 72% des étudiants utilisent la musique comme outil de gestion du stress, tout en devant concilier exigences académiques et consommation médiatique intensive.

2. Méthodologie

a. Population d'étude

La population cible est constituée d'étudiants de l'enseignement supérieur français, sans restriction disciplinaire. Ce choix se justifie par :

- ♪ Une exposition accrue aux pratiques numériques d'écoute
- ♪ Des rythmes de vie susceptibles d'influencer les habitudes (cours, temps libre, transports)
- ♪ Une homogénéité relative en termes de tranche d'âge (18-25 ans majoritairement)

b. Recueil des données

Méthode d'échantillonnage :

- 🎧 Échantillonnage non probabiliste par convenance, via des réseaux étudiants
- 🎧 Biais potentiel : surreprésentation des étudiants actifs sur les groupes en ligne.

Instrument :

- 🎧 Questionnaire auto-administré sur Google Forms (avantages : accessibilité, traitement automatisé)
- 🎧 Questions mêlant échelles de Likert (ex : "À quelle fréquence écoutez-vous de la musique pendant les trajets ?") et questions ouvertes

Diffusion :

- 🎧 Canaux principaux : groupes Facebook étudiants, fils WhatsApp universitaires
- 🎧 Durée limitée à 7 jours pour concentrer les réponses et éviter la dilution temporelle

Période d'enquête :

- 🎧 Collecte réalisée sur deux semaines (hors vacances académiques) pour capter des comportements habituels

c. Descriptif des variables

Nom	Libellé explicite	Valeurs / Unité	Nature
sexe	Sexe de l'étudiant	1 = Homme, 2 = Femme, Autre	Qualitative nominale
genre_preferé	Genre musical préféré	1 = Rap, 2 = Pop, 3 = Rock, 4 = Electro, 5 = Classique, 6 = Jazz, 7=Autre	Qualitative nominale
utilisation_musique	Utilisation principale de la musique	1 = Pour faire du sport 2 = Pour me concentrer / travailler 3 = Pour me détendre 4 = Pendant les trajets	Qualitative nominale
support_ecoute	Support d'écoute privilégié	1 = Spotify / Deezer / Apple Music 2 = Vinyle/CD 3 = YouTube	Qualitative nominale
moment_ecoute	Moment préféré pour écouter de la musique	1 = Le matin 2 = L'après-midi 3 = Le soir	Qualitative ordinale
frequence_ecoute	Fréquence d'écoute	1 = Tous les jours 2 = Presque tous les jours 3 = Quelques fois par semaine	Qualitative ordinale
age	Âge de l'étudiant	Entier en années (ex : 18, 22, 27)	Quantitative discrète
stress	Niveau de stress perçu	Échelle de 1 à 10	Quantitative discrète
duree_ecoute	Nombre d'heures d'écoute musicale moyenne par jour	Nombre d'heures (ex : 2,5, 1,75, 0,5)	Quantitative continue
temps_ecran	Nombres d'heures passées en moyenne par jour sur un écran	Nombre d'heures (ex : 4, 7)	Quantitative discrète

3. Caractéristiques de l'échantillon

Notre échantillon est de taille $n = 78$, il est composé à 100% d'étudiants français. Notre échantillon (1h30/jour) présente une durée d'écoute inférieure à la moyenne nationale (2h11/jour) et nettement en dessous des jeunes de moins de 25 ans (2h43/jour). Cette différence pourrait s'expliquer par des spécificités méthodologiques (critères de sélection des répondants) ou contextuelles (usage simultané d'écrans pour d'autres activités). Les données nationales indiquent que 94 % des Français écoutent quotidiennement de la musique, avec des pics en fin de journée et une préférence marquée pour les langues française et anglaise.

4. Analyse univariée (voir toutes les variables dans Annexe 1 page 16)

a. Variable qualitative nominale

La variable **sexe** est une variable qualitative nominale qui permet de distinguer les individus selon leur genre. Elle prend deux modalités : **Homme (H)** et **Femme (F)**.

1) Tableau de distribution

Sexe	Fréquence relative
Homme (H)	0,564 (56,4 %)
Femme (F)	0,436 (43,6 %)

Ce tableau montre que les hommes représentent **56,4 %** de l'effectif total, tandis que les femmes en constituent **43,6 %**.

2) Visualisation graphique

```
barplot(prop.table(table(sexe)),
        main="Distribution du sexe",
        ylab="Fréquence", xlab="Sexe", ylim=c(0,1),
        col="cyan")
```

Un diagramme en colonnes a été réalisé afin d'illustrer la répartition des individus selon le sexe. Il met en évidence une prédominance des hommes = 1 par rapport aux femmes = 2 dans l'échantillon.



La distribution de la variable **sexe** indique une légère surreprésentation des hommes dans l'échantillon. Cette répartition peut avoir un impact sur certaines analyses ultérieures, notamment si des comparaisons entre genres sont effectuées.

b. Variable qualitative ordinale

La variable **frequence_ecoute** est une variable qualitative ordinale, car ses modalités peuvent être classées selon un ordre croissant d'habitude d'écoute. Elle permet de mesurer la régularité d'écoute des individus.

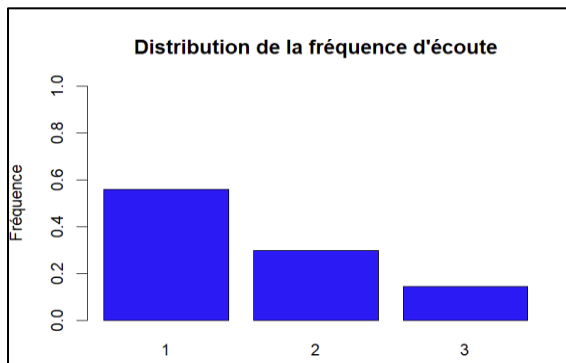
a) Tableau de distribution

Fréquence d'écoute	Fréquence relative
Tous les jours	0,558 (55,8 %)
Presque tous les jours	0,299 (29,9 %)
Quelques fois par semaine	0,143 (14,3 %)

Ce tableau montre que la répartition de la population pour la variable de la fréquence d'écoute.

b) Visualisation graphique

```
barplot(prop.table(table(frequence_ecoute)))
barplot(prop.table(table(frequence_ecoute)),
        main="Distribution de la fréquence d'écoute",
        ylab="Fréquence", xlab="fréquence d'écoute", ylim=c(0,1),
        col="blue")
```



Le diagramme en colonnes illustre la répartition des individus selon leur fréquence d'écoute. Les hauteurs des barres représentent la proportion d'individus pour chaque modalité.

On observe que la barre correspondant à "**Tous les jours**" = 1 est la plus élevée, traduisant une écoute quotidienne majoritaire. À l'inverse, la modalité "**Quelques fois par semaine**" = 3 est la moins représentée.

c) Interprétation

Les résultats mettent en évidence une tendance marquée vers une écoute régulière, avec une forte majorité d'individus écoutant tous les jours (55,8 %) ou presque tous les jours (29,9 %). Seuls 14,3 % des individus écoutent plus occasionnellement, quelques fois par semaine.

Cette distribution suggère que l'échantillon est composé en grande partie d'auditeurs journaliers, ce qui peut influencer les analyses futures.

c. Variable quantitative continue

duree_ecoute

```
> summary(duree_ecoute)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.250   0.750   1.075   1.395   2.000   4.000
```

Statistiques principales

- ♪ Durée minimale d'écoute : 0,25 heure (15 minutes)
- ♪ 1er quartile (Q1) : 0,75 heure (45 minutes) → 25 % des individus écoutent **moins de 45 min** de musique par jour.
- ♪ Médiane (Q2) : 1,075 heure (1h04) → 50 % des individus écoutent **moins de 1h04** de musique par jour, et 50 % écoutent plus.
- ♪ 3e quartile (Q3) : 2 heures → 75 % des individus écoutent **moins de 2h** de musique par jour.
- ♪ Durée maximale d'écoute : 4 heures

Moyenne et dispersion

- 🎧 **Moyenne** : 1,395 heure → En moyenne, les individus écoutent environ **1h24 min** de musique par jour.

```
> sd(duree_ecoute)
[1] 0.8354014
```

- 🎧 **Écart-type élevé** : 0,83 → Les valeurs sont assez dispersées autour de la moyenne.

```
> sd(duree_ecoute)/mean(duree_ecoute) # coef de variation
[1] 0.5989091
```

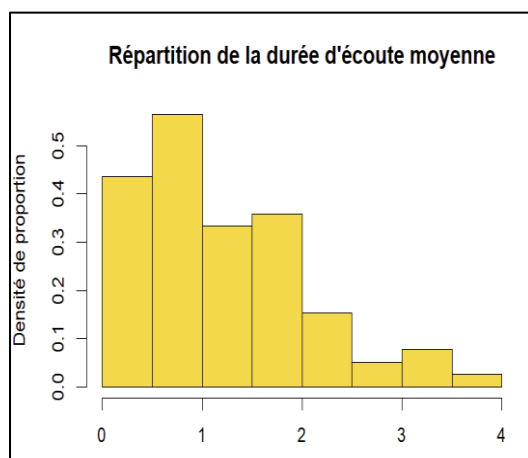
- 🎧 **Coefficient de variation** : 0,5989 → La variabilité est **modérée à élevée**, ce qui signifie que les habitudes d'écoute varient beaucoup d'un individu à l'autre.

Centiles et répartition des durées

```
> quantile(duree_ecoute, probs=seq(0.01, 1, by=0.01))
```

- 🎵 10 % des individus écoutent au maximum 30 minutes à 1 heure.
- 🎵 50 % des individus écoutent entre 1h et 2h de musique par jour.
- 🎵 Seulement 10 % des individus dépassent 2h50 d'écoute quotidienne.

Analyse de l'histogramme



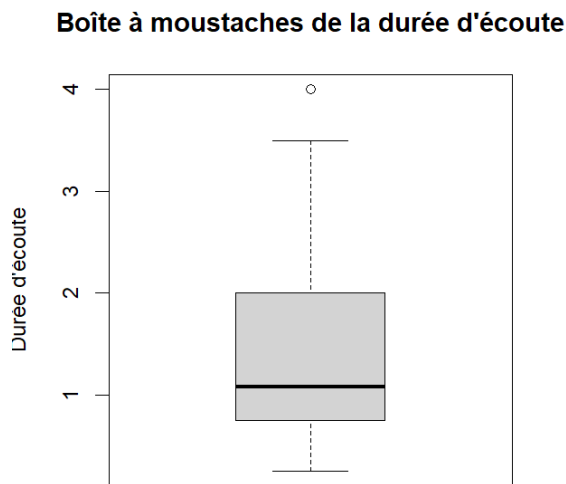
Distribution: La distribution est asymétrique à droite. Cela signifie que la majorité des étudiants ont une durée d'écoute moyenne plus courte, avec une diminution progressive du nombre d'étudiants à mesure que la durée d'écoute augmente.

Fréquence: La répartition nous montre que la majorité des étudiants écoutent de la musique majoritairement entre 0 et 2 heures. La hauteur des barres diminue ensuite progressivement, montrant que moins d'étudiants écoutent pendant des durées plus longues.

Boîte à moustaches

```
> boxplot(duree_ecoute, main="Boîte à moustaches de l'âge", ylab="Durée d'écoute")  
> quantile(duree_ecoute, probs=seq(0.01, 1, by=0.01))
```

La durée maximale d'écoute est 4h, ce qui dépasse la borne supérieure de 3,875h. Il y a donc une valeur atypique supérieure à 3,875h, correspondant au plus gros consommateur de musique.



La boîte à moustache est définie par :

- 🔊 Q1 (1er quartile) = 0,75 h
- 🔊 Q3 (3e quartile) = 2 h
- 🔊 IQR (Intervalle interquartile) = $Q3 - Q1 = 2 - 0,75 = 1,25$ h

Les bornes pour détecter les valeurs atypiques sont :

- 🔊 Borne inférieure = $Q1 - 1.5 \times IQR = 0,75 - (1.5 \times 1,25) = -1,125$ h (impossible, donc la borne inférieure sera 0 h)
- 🔊 Borne supérieure = $Q3 + 1.5 \times IQR = 2 + (1.5 \times 1,25) = 3,875$ h

Dans la partie centrale de la distribution, la médiane est plus proche de Q1 que de Q3, indiquant une répartition asymétrique. Cela signifie que la majorité des individus écoutent moins que la moyenne (1h30), mais quelques-uns écoutent bien plus longtemps.

En queue de distribution :

La moustache supérieure ($Q3 + 1.5 \times IQ$) est beaucoup plus étendue que la moustache inférieure, confirmant une asymétrie à droite.

Présence de valeurs atypiques au-delà de 3h, qui allongent la queue de distribution vers la droite.

d. Variable quantitative discrète

> summary(age)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
18.00	20.00	21.00	21.22	22.00	26.00

- 🔊 Minimum = 18 ans
- 🔊 1er quartile (Q1) = 20 ans → 25 % des individus ont **moins de 20 ans**.
- 🔊 Médiane (Q2) = 21 ans → 50 % des individus ont **moins de 21 ans**.
- 🔊 Moyenne = 21,22 ans
- 🔊 3e quartile (Q3) = 22 ans → 75 % des individus ont **moins de 22 ans**.
- 🔊 Maximum = 26 ans

```
> sd(age)
[1] 1.770226
```

- 🔊 **Écart-type élevé** : 1,77 → Les valeurs sont assez dispersées autour de la moyenne.

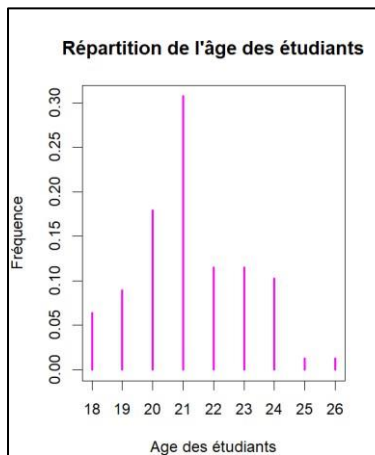
```
> sd(age)/mean(age)
[1] 0.08343058
```

- 🔊 **Coefficient de variation** = 0,0834 → Cela signifie que l'âge des individus est peu dispersé autour de la moyenne.

- 🔊 IQR (Intervalle interquartile) = $Q3 - Q1 = 22 - 20 = 2$ ans

Analyse du diagramme en bâtons

```
plot(prop.table(table(age)),type="h",col="magenta",ylab="Fréquence",
     xlab="Age des étudiants",
     main="Répartition de l'âge des étudiants")
```



Distribution

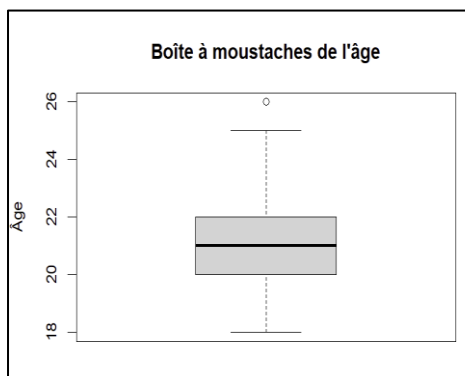
La distribution de l'âge des étudiants est unimodale, avec un pic clair autour de 21 ans. Elle est légèrement asymétrique à droite, indiquant que la majorité des étudiants ont autour de 21 ans, mais qu'il y a quelques étudiants plus âgés.

Fréquence

La barre la plus haute correspond à l'âge de 21 ans, ce qui signifie que c'est l'âge le plus fréquemment observé dans cet échantillon d'étudiants. Les barres diminuent progressivement de 21 à 26 ans, montrant que moins d'étudiants ont ces âges plus élevés.

Boîte à moustaches

```
boxplot(age, main="Boîte à moustaches de l'âge",
        ylab="Âge")
```



```
> quantile(age, probs=seq(0.01,1,by=0.01)) # centiles
```

Les bornes pour détecter les valeurs atypiques sont :

- Borne inférieure = $Q1 - 1.5 \times IQR = 20 - (1.5 \times 2) = 17$ ans (impossible, donc la borne inférieure reste 18 ans, qui est la valeur minimale)
- Borne supérieure = $Q3 + 1.5 \times IQR = 22 + (1.5 \times 2) = 25$ ans

La valeur maximale de l'âge est d'environ 26 ans, ce qui dépasse la borne supérieure du diagramme. Il y a donc une valeur atypique supérieure, correspondant à l'étudiant le plus âgé.

Dans la partie centrale de la distribution, la médiane est aussi loin du premier quartile (Q1) que du troisième quartile (Q3), indiquant une répartition relativement symétrique. Cela signifie que la majorité des âges sont regroupés autour de 21 ans avec une dispersion modérée.

En queue de distribution :

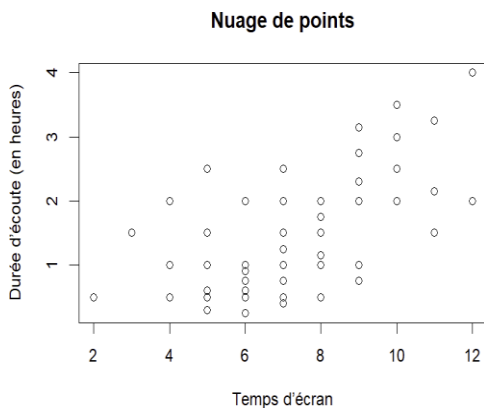
- ♪ La moustache supérieure ($Q3 + 1.5 \times IQR$) est plus étendue que la moustache inférieure, confirmant une asymétrie à droite.
- ♪ Présence d'une valeur atypique au-delà de 25 ans, qui allonge la queue de distribution vers la droite.

5. Analyse bivariée (voir toutes les variables Annexe 2 page 23)

Nous allons réaliser l'analyse univariée du temps d'écran et de la durée d'écoute.

1. Description du nuage de points

```
plot(temps_ecran,duree_ecoute,main="Nuage de points", xlab="Temps d'écran",  
      ylab="Durée d'écoute (en heures)")
```



Le graphique montre un nuage de points représentant la relation entre le temps d'écran (en heures) sur l'axe des abscisses et la durée d'écoute (en heures) sur l'axe des ordonnées.

On observe une tendance générale où les points semblent s'aligner dans une direction ascendante : à mesure que le temps d'écran augmente, la durée d'écoute tend également à augmenter.

Cela suggère une corrélation positive entre les deux variables. Cependant, la dispersion des points indique qu'il ne s'agit pas d'une relation parfaitement linéaire.

2. Test de corrélation linéaire

Le test de corrélation linéaire a été effectué avec les hypothèses suivantes :

- 🎧 H_0 : La durée d'écoute et le temps d'écran ne sont pas corrélés linéairement.
- 🎧 H_1 : La durée d'écoute et le temps d'écran sont corrélés linéairement.

Résultats du test :

```
> cor.test(temps_ecran,duree_ecoute)
```

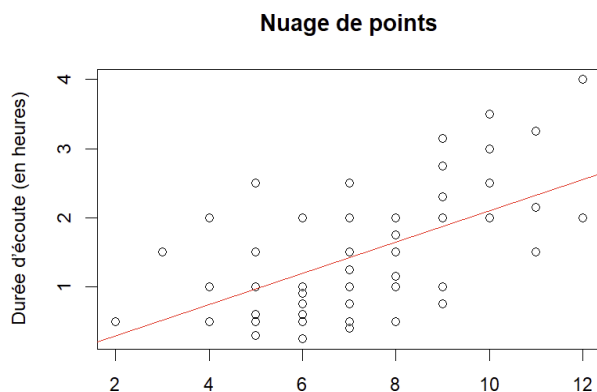
```
Pearson's product-moment correlation  
  
data: temps_ecran and duree_ecoute  
t = 6.1996, df = 76, p-value = 2.706e-08  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
 0.4098662 0.7104464  
sample estimates:  
 cor  
0.5795385
```

- 🎵 Statistique t : 6.1996
- 🎵 Degrés de liberté (df) : 76
- 🎵 p-value : 2.706e-08 (extrêmement faible)
- 🎵 Intervalle de confiance à 95 % : [0.4098662 ; 0.7104464]
- 🎵 Coefficient de corrélation (r) : 0.5795385

Interprétation :

1. Coefficient de corrélation linéaire ($r = 0.579$) :
 - 🎵 Le coefficient de corrélation est positif et modéré, indiquant une relation linéaire significative entre le temps d'écran et la durée d'écoute.
 - 🎵 Cela signifie qu'une augmentation du temps d'écran est associée à une augmentation de la durée d'écoute.
2. p-value (2.706e-08) :
 - 🎵 La p-value est très faible, bien inférieure au seuil classique de 0.05.
 - 🎵 On rejette donc l'hypothèse nulle (H_0) et accepte l'hypothèse alternative (H_1) : il existe une corrélation linéaire significative entre les deux variables.
3. Intervalle de confiance :
 - 🎵 L'intervalle de confiance pour le coefficient de corrélation est compris entre 0.41 et 0.71, ce qui confirme que la relation est statistiquement significative.

Il existe une corrélation linéaire modérée et significative entre le temps d'écran et la durée d'écoute ($r = 0.58$, $p < 0.001$). Les résultats suggèrent que ces deux variables sont liées : plus le temps passé devant un écran augmente, plus la durée d'écoute tend à augmenter également.



Si on fait un test de régression linéaire avec comme résultat obtenu :

Coefficients:

(Intercept) temps_ecran
-0.1554 0.2252

Le nuage de points suggère une faible tendance positive entre le temps d'écran et la durée d'écoute. Avec une pente de 0.2252, chaque heure supplémentaire d'écran est associée à une augmentation d'environ 13.5 minutes d'écoute. L'ordonnée à l'origine de -0.1554, bien que non interprétable directement, indique qu'il faut évaluer l'adéquation globale du modèle linéaire. L'équation de la droite est : $duree_ecoute = -0.1554 + 0.2252 * \text{temps_ecran}$

Nous allons faire l'analyse bivariée de la durée d'écoute et de la fréquence d'écoute

Résumés numériques des distributions conditionnelles

```
tapply(duree_ecoute, frequence_ecoute, summary)
```

- 🔊 **Tous les jours** : moyenne = 1.67, médiane = 1.50, min = 0.50, max = 4.00, variance = 0.54
- 🔊 **Presque tous les jours** : moyenne = 1.27, médiane = 0.75, min = 0.25, max = 3.15, variance = 0.91
- 🔊 **Quelques fois par semaine** : moyenne = 0.64, médiane = 0.50, min = 0.30, max = 1.50, variance = 0.11

Ces résultats montrent que les auditeurs quotidiens ont une durée d'écoute moyenne plus élevée et une dispersion relativement faible par rapport aux autres groupes.

Rapport de corrélation (η^2) Le rapport de corrélation (η^2) mesure la part de variance expliquée par la variable qualitative fréquence d'écoute. Voici les étapes du calcul :

```
> var(duree_ecoute)
[1] 0.6978954
🔊 Variance totale : Var(durée d'écoute) = 0.6979
> varintra=sum(table(frequence_ecoute)*varcond)/78
> varintra
[1] 0.5817378
🔊 Variance intra-groupes :0.5817
> varinter=sum(table(frequence_ecoute)*(mcond-mean(duree_ecoute))^2)/78
> varinter
[1] 0.1252207
🔊 Variance inter-groupes :0.1252
> varinter/var(duree_ecoute)
[1] 0.1794262
🔊 Rapport  $\eta^2$  : 0.1252/0.6979=0.179
```

Interprétation : La valeur de $\eta^2 = 0.1794$ indique une liaison modérée entre la fréquence et la durée d'écoute (avec comme seuil de référence 0.2).

Test de comparaison ANOVA

```
> leveneTest(duree_ecoute, frequence_ecoute)
Levene's Test for Homogeneity of Variance (center = median)
  Df F value Pr(>F)
group 2 3.5234 0.03454 *
    74
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> # étape 2 : test ANOVA de comparaison des moyennes (version pour variances
> # inégales grâce à l'argument var.equal=FALSE)
> oneway.test(duree_ecoute~frequence_ecoute, var.equal=FALSE)

One-way analysis of means (not assuming equal variances)

data: duree_ecoute and frequence_ecoute
F = 22.793, num df = 2.000, denom df = 37.852, p-value = 3.185e-07
```

Étape 1 : Test de Levene (égalité des variances)

- 🔊 H_0 : Les variances des groupes sont égales.
- 🔊 H_1 : Les variances des groupes sont différentes.
- 🔊 **Résultat** : p-value = 0.0345 (< 0.05)
- 🔊 **Conclusion** : On rejette H_0 . Les variances sont inégales entre les groupes.

Étape 2 : Test ANOVA (moyennes avec variances inégales)

- 🔊 H_0 : Les moyennes des groupes sont égales.
- 🔊 H_1 : Au moins deux groupes ont des moyennes différentes.
- 🔊 **Résultat** : $F(2,37.85) = 22.79$, p-value = $3,185 \times 10^{-7}$
- 🔊 **Conclusion** : On rejette H_0 . Il existe une différence significative entre les moyennes.

Comparaisons multiples

```
> pairwise.t.test(duree_ecoute, frequence_ecoute, p.adjust.method= "none", pool.sd=FALSE)

Pairwise comparisons using t tests with non-pooled SD

data: duree_ecoute and frequence_ecoute

  1      2
2 0.0895 -
3 6.3e-08 0.0086

P value adjustment method: none
```

Les tests de comparaisons par paires montrent quelles moyennes diffèrent :

- **Tous les jours vs Presque tous les jours** : p-value = 0.0895 (non significatif)
- **Tous les jours vs Quelques fois/semaine** : p-value < 10^{-7} (très significatif)
- **Presque tous les jours vs Quelques fois/semaine** : p-value = 0.0086 (significatif)

Interprétation : Les auditeurs quotidiens et presque quotidiens ont des durées similaires, mais ils diffèrent fortement des auditeurs occasionnels.

La fréquence d'écoute influence significativement la durée d'écoute avec une liaison modérée ($\eta^2 = 0.1794$). Les auditeurs occasionnels se distinguent nettement des autres groupes en ayant une durée moyenne beaucoup plus faible et une dispersion réduite.

! **Nous avons trouvé une analyse bivariée pertinente entre deux variables qualitatives, celle-ci se trouve en annexe page 35!**

6. Régression linéaire multiple

Modèle initial : sélection du modèle pas-à-pas descendante

```
Call:
lm(formula = duree_ecoute ~ frequence_ecoute + temps_ecran +
    stress + genre_prefere + support_ecoute)

Residuals:
    Min       1Q   Median       3Q      Max
-1.1412 -0.3414 -0.1276  0.2059  1.6665

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.08999    0.38884   2.803 0.006519 **
frequence_ecoute -0.08493    0.12111  -0.701 0.485427
temps_ecran    0.17900    0.03514   5.094 2.77e-06 ***
stress        -0.15445    0.03826  -4.037 0.000135 ***
genre_prefere  -0.02155    0.05388  -0.400 0.690383
support_ecoute  0.04254    0.10023   0.424 0.672565
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6003 on 71 degrees of freedom
(1 observation effacée parce que manquante)
Multiple R-squared:  0.5225,    Adjusted R-squared:  0.4888
F-statistic: 15.54 on 5 and 71 DF,  p-value: 2.642e-10
```

Commencement avec le modèle `mod1 = lm(duree_ecoute ~ frequence_ecoute + temps_ecran + stress + genre_prefere + support_ecoute)`. Ce modèle comprend toutes les variables significatives à 10% identifiées lors des analyses bivariées.

$$duree_ecoute_1 = \beta_0 + \beta_1 \cdot frequence_ecoute_1 + \beta_2 \cdot temps_ecran_1 + \beta_3 \cdot stress_1 + \beta_4 \cdot genre_prefere_1 + \beta_5 \cdot support_ecoute_1 + \epsilon_1$$

$$duree_ecoute_1 = 1.08999 - 0.08493 \cdot frequence_ecoute_1 + 0.17900 \cdot temps_ecran_1 - 0.15445 \cdot stress_1 - 0.02155 \cdot genre_prefere_1 + 0.04254 \cdot support_ecoute_1 + \epsilon_1$$

- Fréquence d'écoute : Toutes choses égales par ailleurs, une augmentation de la fréquence d'écoute de musique (par exemple, passer de "quelques fois par semaine" à "presque tous les jours") est associée à une diminution de la durée d'écoute de 0.08493 heure (environ 5 minutes), mais cette relation n'est pas significative ($p > 0.05$).
- Temps d'écran : Toutes choses égales par ailleurs, chaque heure supplémentaire passée devant un écran est associée à une augmentation significative de la durée d'écoute de musique de 0.179 heure (environ 10 minutes) ($p < 0.001$).
- Stress : Toutes choses égales par ailleurs, chaque point supplémentaire sur l'échelle de stress est associé à une diminution significative de la durée d'écoute de musique de 0.15445 heure (environ 9 minutes) ($p < 0.001$).
- Genre préféré : Toutes choses égales par ailleurs, le genre musical préféré a un impact négatif de -0.02155 heure (environ 1 minute) sur la durée d'écoute, mais cet effet n'est pas significatif ($p > 0.05$).
- Support d'écoute : Toutes choses égales par ailleurs, le type de support utilisé entraîne une variation positive de la durée d'écoute de 0.04254 heure (environ 3 minutes), mais cet effet n'est pas significatif ($p > 0.05$).

♪ **Étape 1 :** Identification de `genre_prefere` comme la variable la moins significative ($p = 0.69$), nous l'avons éliminée en première.

♪ **Étape 2 :** Ensuite, `support_ecoute` (p -value = 0.67) car elle était la moins significative parmi les variables restantes.

🎵 **Étape 3 :** Enfin, `frequence_ecoute` (p-value = 0.42).

Modèle final : Aboutissement du modèle final `mod1 = lm(duree_ecoute ~ temps_ecran + stress)`. Dans ce modèle, les deux variables restantes sont significatives à un seuil de 5% (`temps_ecran` $p = 8.98e-07$, `stress` $p = 4.20e-06$).

```
> mod1 = lm(duree_ecoute ~ temps_ecran + stress)
> summary(mod1)

Call:
lm(formula = duree_ecoute ~ temps_ecran + stress)

Residuals:
    Min       1Q   Median       3Q      Max
-1.20074 -0.31076 -0.09572  0.23429  1.78165

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.02307    0.32959   3.104  0.00269 **
temps_ecran  0.17751    0.03314   5.356  8.98e-07 ***
stress      -0.16748    0.03374  -4.965  4.20e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5985 on 75 degrees of freedom
Multiple R-squared:  0.5001,    Adjusted R-squared:  0.4868
F-statistic: 37.52 on 2 and 75 DF,  p-value: 5.094e-12
```

$$\text{duree_ecoute}_i = \beta_0 + \beta_1 \cdot \text{temps_ecran}_i + \beta_2 \cdot \text{stress}_i + \epsilon_i$$
$$\text{duree_ecoute} = 1.02307 + 0.17751 \cdot \text{temps_ecran} - 0.16748 \cdot \text{stress} + \epsilon$$

Temps d'écran : Toutes choses égales par ailleurs, une heure supplémentaire passée sur un écran est associée à une augmentation significative de la durée d'écoute de musique de 0.17751 heure (environ 10.65 minutes).

Stress : Toutes choses égales par ailleurs, une augmentation d'un point du niveau de stress est associée à une diminution significative de la durée d'écoute de musique de 0.16748 heure (environ 10 minutes) .

Hypothèses :

- 🎧 H_0 : Le modèle avec `temps_ecran` et `stress` n'explique pas significativement la variance de `duree_ecoute`.
- 🎧 H_1 : Au moins l'un des coefficients de `temps_ecran` ou `stress` est différent de zéro, indiquant que le modèle explique significativement la variance de `duree_ecoute`.

Vérification : La p-value du test F ($5.094e-12$) est inférieure à 0.05. On rejette donc l'hypothèse nulle.

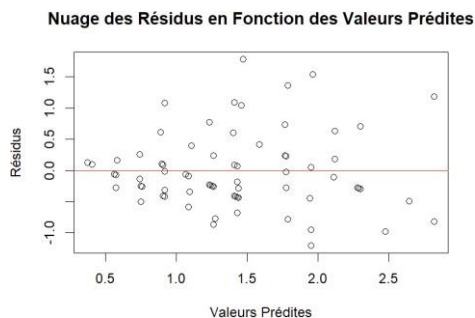
Le modèle avec `temps_ecran` et `stress` explique significativement la variance de `duree_ecoute`.

Etude des résidus

Le R-squared ajusté est de 0.4868. Cela signifie que 48.68% de la variance de la durée d'écoute de musique est expliquée par le modèle, c'est-à-dire par le temps passé sur un écran et le niveau de stress, en ajustant pour la complexité du modèle.

En regardant le Q-Q plot des résidus, on peut observer les points suivants :

- ♪ **Central** : La plupart des points se situent relativement près de la ligne droite, ce qui indique que la distribution des résidus est raisonnablement normale au centre.

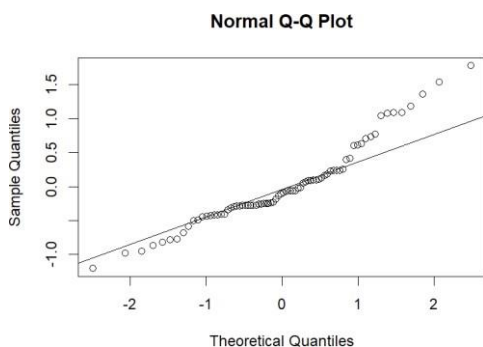


- ♪ **Extrémities** : On observe une déviation des points par rapport à la ligne droite aux extrémités du graphique. Cela suggère des écarts à la normalité dans les queues de distribution.

Interprétation :

Bien que la partie centrale des résidus semble suivre une distribution normale, les déviations observées aux extrémités suggèrent que la distribution des résidus n'est pas parfaitement normale. Ces écarts peuvent indiquer la présence de quelques valeurs extrêmes ou une légère asymétrie dans la distribution des résidus.

L'hypothèse de normalité des résidus est approximativement respectée, mais il existe des écarts notables aux extrémités. Il faut considérer cet aspect dans l'interprétation globale du modèle. Une légère non-normalité n'invalidé pas nécessairement le modèle, surtout si les autres hypothèses sont respectées. Cependant, si les écarts sont importants, il pourrait être utile d'explorer des transformations de variables ou d'autres approches de modélisation.



En se basant sur le nuage des résidus en fonction des valeurs prédites, voici une interprétation du nuage des résidus en fonction des valeurs prédites :

- ♪ **Répartition aléatoire** : Les points semblent être répartis de manière aléatoire autour de la ligne horizontale rouge (qui représente zéro). Il n'y a pas de motif évident, comme une forme de cône (qui indiquerait une hétéroscédasticité) ou une courbure (qui indiquerait une non-linéarité).
- ♪ **Homoscédasticité** : La variance des résidus semble relativement constante sur toute la plage des valeurs prédites. Il n'y a pas d'élargissement ou de rétrécissement notable du nuage de points.
- ♪ **Linéarité** : L'absence de motif courbé suggère que la relation entre les variables explicatives et la variable dépendante est raisonnablement linéaire.

Conclusion : Le nuage des résidus en fonction des valeurs prédites ne révèle pas de problèmes majeurs. L'hypothèse d'homoscédasticité semble respectée, et il n'y a pas d'indication de non-linéarité. Cela renforce la validité de ton modèle linéaire.

```
> res.std <- mod$residuals / 0.5985
> mod$residuals/0.5985
 1      2      3      4      5      6
1.23108026 2.28361408 -0.67792139 0.43733476 1.17672743 0.11990499
 7      8      9     10     11     12
1.74047681 1.97455381 -1.13322784 2.97685232 -0.13909107 -0.73227422
 13     14     15     16     17     18
-0.57355933 -0.97451296 -0.49411635 -1.36713375 -0.45652084 -0.71551690
 19     20     21     22     23     24
-2.00624524 -1.62612981 -0.45652084 0.16158137 0.39565837 0.17425782
 25     26     27     28     29     30
1.82426341 -0.10149556 1.02643703 1.82426341 1.28543309 1.00559884
 31     32     33     34     35     36
-0.10557643 -0.10149556 -0.45652084 -1.58853430 -0.29780595 -0.41484445
 37     38     39     40     41     42
-1.28786185 -0.83255540 -0.41892532 -0.69467871 -0.43160177 1.81158696
 43     44     45     46     47     48
-1.30870004 0.16158137 0.21185333 0.28270075 -0.41484445 -0.10149556
 49     50     51     52     53     54
0.66281618 -0.45652084 -0.11825288 -0.43568264 0.08230948 -0.73635509
 55     56     57     58     59     60
0.15341963 -0.18076745 0.39565837 0.39973924 0.37890105 -0.40216801
 61     62     63     64     65     66
-0.46060171 -0.47735903 0.69224994 -0.68200226 -0.38541069 0.15750050
 67     68     69     70     71     72
-0.69875958 -1.43818891 -0.45242163 -0.23100276 -0.02634120 2.57181782
 73     74     75     76     77     78
1.05560806 -0.03880989 -0.52759433 0.30372836 -0.81991561 -0.48164765
> outliers <- which(abs(res.std) > 2)
> print(outliers)
 2 10 19 72
 2 10 19 72
```

L'analyse des résidus standardisés révèle la présence de quatre observations (2, 10, 19 et 72) avec des résidus standardisés supérieurs à 2 en valeur absolue. Plus précisément :

- 🔊 **Observation 2 :** Résidu standardisé de 2.283
- 🔊 **Observation 10 :** Résidu standardisé de 2.976
- 🔊 **Observation 19 :** Résidu standardisé de -2.006
- 🔊 **Observation 72 :** Résidu standardisé de 2.571

Ces valeurs indiquent que ces observations s'écartent le plus du modèle par rapport aux autres, avec l'observation 10 s'écartant le plus. Bien que le modèle soit globalement significatif, il est important de reconnaître que ces observations ne sont pas bien expliquées par le modèle actuel.

7. Conclusion

L'analyse univariée révèle une population étudiante homogène en âge avec des temps d'écran concentrés entre 5h et 8h/jour, malgré quelques valeurs extrêmes. L'analyse bivariable identifie une corrélation modérée entre temps d'écran et écoute musicale, tandis que le stress montre un impact négatif significatif. Aucun lien n'apparaît entre âge, sexe ou contexte d'utilisation et la durée d'écoute, mais les utilisateurs de Spotify/Deezer écoutent significativement plus que ceux de YouTube. Le modèle de régression confirme le rôle clé du temps d'écran et du stress pour expliquer la durée d'écoute, bien que certains *outliers* suggèrent des comportements atypiques non captés par le modèle. En synthèse, les habitudes musicales des étudiants semblent davantage influencées par des facteurs comportementaux (exposition aux écrans, gestion du stress).

ANNEXES :

Annexe 1 : Analyse univariée

a. Variable qualitative nominale

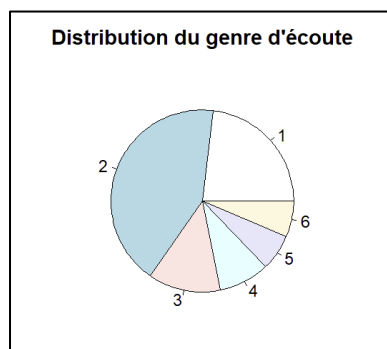
Variable genre préféré

La variable **genre_preferé** est une variable **qualitative nominale**, car elle correspond à des catégories musicales sans ordre particulier. Elle permet d'identifier les préférences musicales des individus.

a) Tableau de distribution

Genre musical préféré	Fréquence relative
Rap	0,231 (23,1 %)
Pop	0,423 (42,3 %)
Rock	0,128 (12,8 %)
Électro	0,090 (9,0 %)
Classique	0,064 (6,4 %)
Jazz	0,064 (6,4 %)

b) Visualisation graphique



Le diagramme en secteurs illustre la répartition des préférences musicales des individus. Chaque portion du cercle représente la part relative de chaque genre musical dans l'échantillon.

On observe que le genre le plus écouté est la Pop =2 (42,3 %), suivi du Rap = 1 (23,1 %). Les autres genres, comme le Rock = 3 (12,8 %), l'Électro =4 (9,0 %), le Classique = 5 (6,4 %) et le Jazz =6 (6,4 %), sont moins représentés.

c) Interprétation

Les résultats indiquent une forte popularité de la Pop (42,3 %) et du Rap (23,1 %), qui rassemblent ensemble plus des deux tiers (65,4 %) des préférences. À l'inverse, les genres Classique et Jazz (6,4 % chacun) sont les moins écoutés par les individus de l'échantillon.

Cette distribution peut être mise en relation avec d'autres variables, comme l'âge ou le moment d'écoute, afin d'identifier des tendances dans les habitudes musicales des individus.

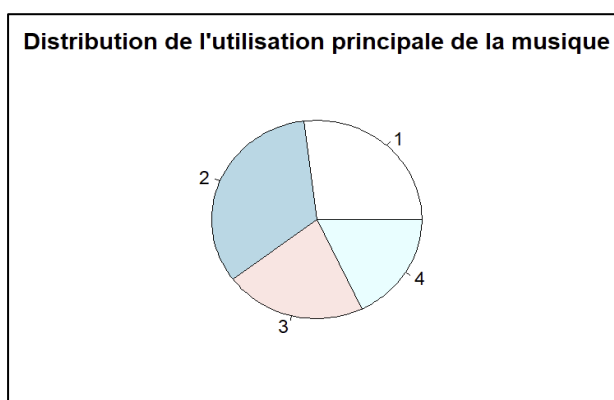
Variable utilisation de la musique

La variable **utilisation_musique** est une variable **qualitative nominale**, car elle représente différentes raisons d'écouter de la musique, sans ordre hiérarchique entre elles. Elle permet d'identifier les principaux contextes d'écoute des individus.

a) Tableau de distribution

Utilisation de la musique	Fréquence relative
Pour faire du sport	0,269 (26,9 %)
Pour me concentrer / travailler	0,333 (33,3 %)
Pour me détendre	0,218 (21,8 %)
Pendant les trajets	0,179 (17,9 %)

b) Visualisation graphique



Le diagramme en secteurs illustre la répartition des différentes utilisations de la musique. Chaque portion du cercle représente la part relative de chaque raison d'écoute.

On observe que l'utilisation principale de la musique est pour se concentrer/travailler = 2 (33,3 %), suivie de l'écoute pendant le sport = 1 (26,9 %). La détente = 3 (21,8 %) et l'écoute pendant les trajets = 4 (17,9 %) sont des motivations légèrement moins fréquentes.

c) Interprétation

Les résultats indiquent que la musique est majoritairement utilisée comme un outil de concentration (33,3 %) et d'accompagnement pour le sport (26,9 %). L'écoute pour la détente (21,8 %) et pendant les trajets (17,9 %) restent également des usages répandus, bien que moins fréquents.

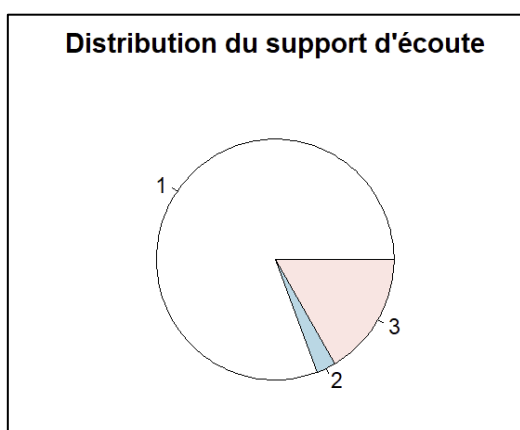
Cette répartition met en avant l'importance fonctionnelle de la musique dans le quotidien des individus.

La variable **support_ecoute** est une variable **qualitative nominale**, car elle représente différents moyens d'accéder à la musique, sans ordre particulier. Elle permet d'identifier les supports privilégiés par les individus pour écouter de la musique.

a) Tableau de distribution

Support d'écoute	Fréquence relative
Spotify / Deezer / Apple Music	0,808 (80,8 %)
Vinyle / CD	0,026 (2,6 %)
YouTube	0,167 (16,7 %)

b) Visualisation graphique



Le diagramme en illustre la répartition des différents supports d'écoute. Chaque portion du cercle représente la part relative de chaque support utilisé.

On observe que l'écrasante majorité des individus (80,8 %) privilégient les plateformes de streaming = 1 (Spotify, Deezer, Apple Music). YouTube = 3 est utilisé par 16,7 % des individus, tandis que les supports physiques = 2 (Vinyle/CD) sont très minoritaires (2,6 %).

c) Interprétation

Ces résultats mettent en évidence la domination des plateformes de streaming (80,8 %), confirmant leur place centrale dans les habitudes d'écoute actuelles. YouTube (16,7 %) reste un acteur non négligeable, souvent utilisé pour l'accès gratuit à la musique, tandis que l'usage des supports physiques (2,6 %) demeure très marginal, probablement réservé aux amateurs de formats traditionnels ou de collections.

b. Variable qualitative ordinale

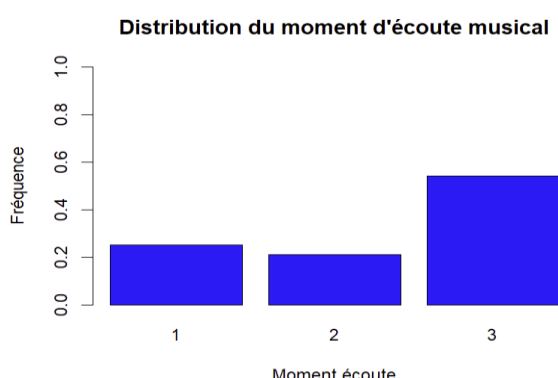
Variable moment d'écoute

La variable **moment_ecoute** est une variable **qualitative ordinale**, car ses modalités suivent un ordre temporel (matin → après-midi → soir). Elle permet d'identifier les moments privilégiés d'écoute des individus.

a) Tableau de distribution

Moment d'écoute	Fréquence relative
Le matin	0,250 (25,0 %)
L'après-midi	0,211 (21,1 %)
Le soir	0,539 (53,9 %)

b) Visualisation graphique



Le diagramme en colonnes ci-dessous illustre la répartition des individus selon leur moment d'écoute privilégié. Chaque barre représente la proportion d'individus pour chaque période de la journée.

On constate que le soir = 1 est le moment le plus fréquenté avec 53,9 % des individus, suivi du matin = 2 (25,0 %) et enfin de l'après-midi = 3 (21,1 %), qui est la période la moins courante.

c) Interprétation

Les résultats montrent une préférence marquée pour l'écoute en soirée (53,9 %), suggérant que les individus privilégient ce moment pour se détendre ou s'informer après leurs obligations quotidiennes. Le matin (25,0 %) et l'après-midi (21,1 %) sont moins sollicités, bien qu'ils restent significatifs.

c. Variable quantitative discrète

Variable temps d'écran

Variable temps_ecran est une **variable quantitative discrète**.

```
> summary(temps_ecran)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 2.000   5.000   7.000   6.885   8.000  12.000
```

Statistiques principales

- ♪ Temps minimum d'écran : 2 heures
- ♪ 1er quartile (Q1) : 5 heures → 25 % des individus passent moins de 5h sur les écrans
- ♪ Médiane (Q2) : 7 heures → 50 % des individus passent moins de 7h sur les écrans
- ♪ 3e quartile (Q3) : 8 heures → 75 % des individus passent moins de 8h sur les écrans
- ♪ Temps maximum d'écran : 12h

```
> sd(temps_ecran)
[1] 2.150088
```

- 🔊 **Écart-type élevé** : 2,15 → Les valeurs sont assez dispersées autour de la moyenne.

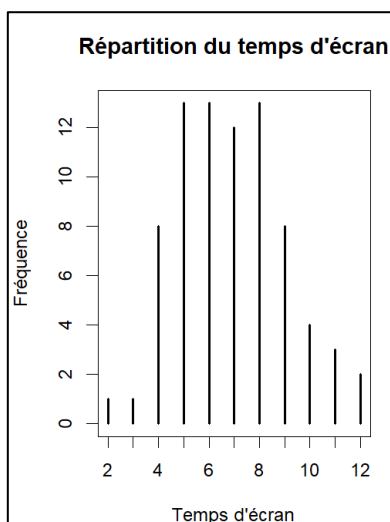
```
> sd(temps_ecran)/mean(temps_ecran)
[1] 0.3123032
```

- 🔊 **Coefficient de variation** = 0,312 → Cela signifie que le temps d'écran des individus est dispersé autour de la moyenne.

- 🔊 **IQR (Intervalle interquartile)** = $Q3 - Q1 = 8 - 5 = 3$ heures

Analyse du diagramme en bâtons

```
plot(table(temps_ecran),
      main = "Répartition du temps d'écran",
      xlab = "Temps d'écran",
      ylab = "Fréquence")
```



Distribution

La distribution du temps d'écran est unimodale, avec un pic clair autour de 5 à 8 heures. Cela indique que la plupart des individus passent environ 5 à 8 heures devant un écran, ce qui correspond à la médiane et à la moyenne.

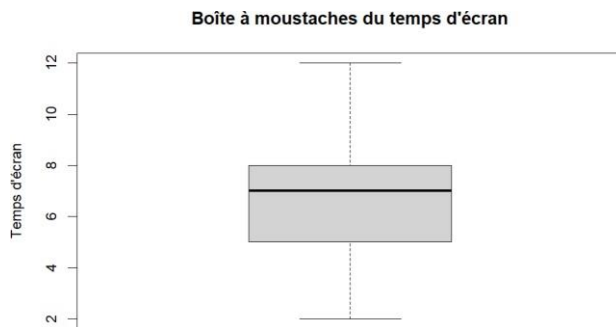
Fréquence

La barre la plus haute correspond aux temps d'écran de 8 heures, ce qui signifie que c'est la durée la plus fréquemment observée. Les fréquences diminuent progressivement pour les temps d'écran plus élevés, montrant qu'un faible nombre d'individus passent plus de 10 heures devant un écran.

Répartition

La distribution est asymétrique à droite, ce qui signifie que bien que la majorité des individus aient un temps d'écran modéré, quelques-uns passent beaucoup plus de temps devant un écran.

Boîte à moustache



La moustache supérieure ($Q3 + 1.5 \cdot IQ$) = $8 + 4.5 = 12.5h$.

La moustache inférieure ($Q1 - 1.5 \cdot IQ$) = $5 - 4.5 = 0.5h$ (mais la valeur min est 2h, donc pas de valeurs trop faibles).

La valeur maximale du temps d'écran est d'environ 12 heures, ce qui ne dépasse pas la borne supérieure du diagramme. Il n'y a donc pas de valeur atypique détectée.

Dans la partie centrale de la distribution, la médiane est plus proche du troisième quartile (Q3) que du premier quartile (Q1), indiquant une légère asymétrie à gauche.

Cela signifie que la majorité des individus ont un temps d'écran plutôt élevé, mais quelques-uns ont une consommation plus faible.

En queue de distribution :

- ♪ La moustache inférieure ($Q1 - 1.5 \cdot IQR$) est plus étendue que la moustache supérieure, confirmant une asymétrie à gauche.
- ♪ Aucune valeur atypique détectée, la distribution est donc relativement homogène.

Variable stress

```
> summary(stress)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000   3.000   5.000   5.077   7.000   9.000
```

Statistiques principales :

- ♪ Niveau de stress minimum : 1
- ♪ 1er quartile (Q1) : 3 → 25 % des individus ressentent moins de 3 de niveau de stress sur 10
- ♪ Médiane (Q2) : 5 → 50 % des individus ressentent moins de 5 de niveau de stress sur 10
- ♪ 3e quartile (Q3) : 7 → 75 % des individus ressentent moins de 7 de niveau de stress sur 10
- ♪ Niveau de stress maximum : 9

```
> sd(stress)
[1] 2.112235
```

- 🔊 **Écart-type élevé** : 2,15 → Les valeurs sont assez dispersées autour de la moyenne.

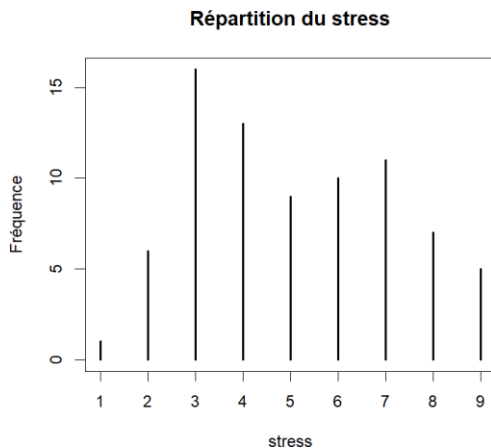
```
> sd(stress)/mean(stress)
[1] 0.4160464
```

🔊 **Coefficient de variation** = 0,416 → Cela signifie que le temps d'écran des individus est dispersé autour de la moyenne.

🔊 **IQR (Intervalle interquartile)** = $Q3 - Q1 = 7 - 3 = 4$ heures

Analyse diagramme en bâtons

```
plot(table(stress),
      main = "Répartition du stress",
      xlab = "stress",
      ylab = "Fréquence")
```



Distribution

La distribution du stress est unimodale, avec un pic clair autour de 3 et 4. Cela indique que la plupart des individus ont un niveau de stress situé autour de ces valeurs, ce qui correspond à la médiane et à la moyenne.

Fréquence

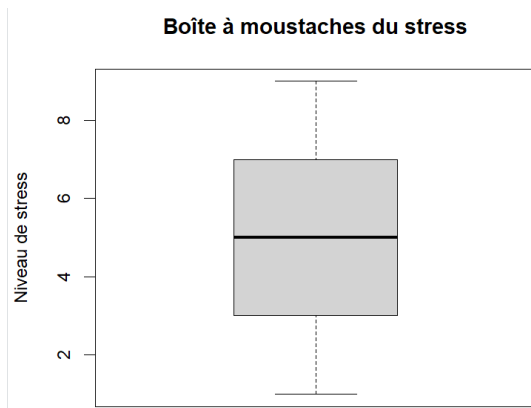
La barre la plus haute correspond au niveau de stress de 3, ce qui signifie que c'est le niveau le plus fréquemment observé. Les fréquences diminuent progressivement pour les niveaux de stress plus élevés, bien que certains individus aient un stress plus important.

Répartition

La distribution semble légèrement asymétrique à droite, ce qui signifie que bien que la majorité des individus aient un niveau de stress modéré, quelques-uns ont des niveaux de stress plus élevés.

Boîte à moustache

```
boxplot(stress, main="Boîte à moustaches du stress",
        ylab="Niveau de stress")
```



La valeur maximale du niveau de stress est d'environ 9, ce qui ne dépasse pas la borne supérieure du diagramme. Il n'y a donc pas de valeur atypique détectée.

Dans la partie centrale de la distribution, la médiane est légèrement plus proche du troisième quartile que du premier quartile, indiquant une légère asymétrie à gauche. Cela signifie que la majorité des individus ont un niveau de stress plutôt élevé, mais quelques-uns ont un stress plus faible. En queue de distribution :

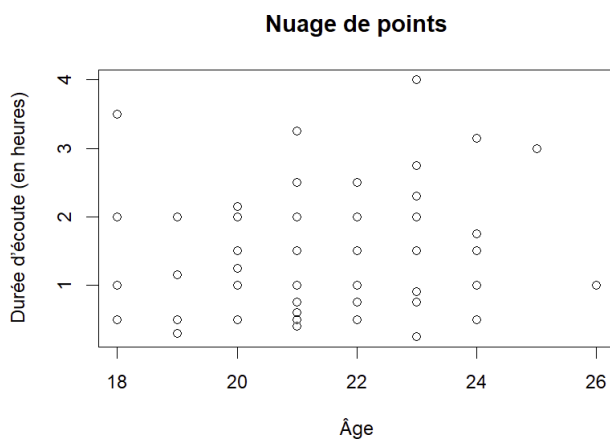
- La moustache inférieure ($Q1 - 1.5 \cdot IQR$) est plus étendue que la moustache supérieure, confirmant une asymétrie à gauche.
- Aucune valeur atypique détectée, la distribution est donc relativement homogène.

6. Analyse bivariable Annexe 2

Analyse bivariable entre âge et durée d'écoute

1. Description du nuage de points

```
> plot(age,duree_ecoute ,main="Nuage de points", xlab="Âge",
+       ylab="Durée d'écoute (en heures)")
```



Le nuage de points montre la distribution des données avec l'âge en abscisse et la durée d'écoute en ordonnée.

La dispersion importante des points suggère une faible relation visuelle entre l'âge et la durée d'écoute. Les points sont éparpillés sans motif clair, indiquant l'absence de tendance où la durée d'écoute augmenterait ou diminuerait avec l'âge.

2. Test de corrélation linéaire

Un test de corrélation de Pearson évalue la relation linéaire entre l'âge et la durée d'écoute.

H_0 : La durée d'écoute et l'âge ne sont pas corrélés linéairement.

H_1 : La durée d'écoute et l'âge sont corrélés linéairement.

```
> cor.test(age,duree_ecoute)

Pearson's product-moment correlation

data:  age and duree_ecoute
t = 1.4623, df = 76, p-value = 0.1478
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.05928823  0.37418139
sample estimates:
cor
0.1654251
```

🎧 Statistique t: 1.4623
 🎧 Degrés de liberté (df): 76
 🎧 p-value: 0.1478
 🎧 Intervalle de confiance à 95 %: $[-0.05928823, 0.37418139]$
 🎧 Coefficient de corrélation (cor): 0.1654251

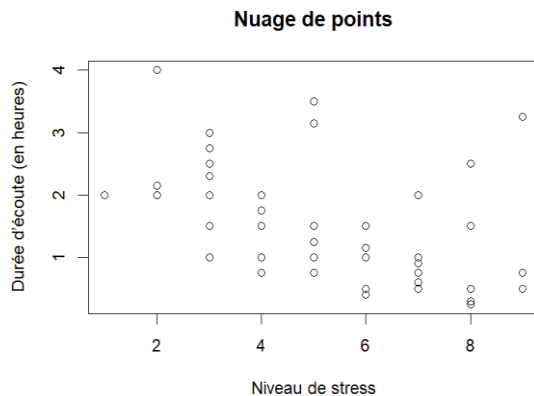
3. Interprétation

Le coefficient de corrélation linéaire ($r = 0.165$) est faible et positif, indiquant une très faible corrélation linéaire entre l'âge et la durée d'écoute. Cela suggère une légère tendance à l'augmentation de la durée d'écoute avec l'âge, mais cette relation est minime.

La p-value (0.1478) est supérieure à 0.05, donc nous ne rejetons pas l'hypothèse nulle.

Il n'existe pas de corrélation linéaire significative entre l'âge et la durée d'écoute ($r = 0.165$, $p > 0.05$). L'âge n'est pas un bon prédicteur de la durée d'écoute, ce que confirme la dispersion observée dans le nuage de points.

Analyse bivariable entre niveau de stress et durée d'écoute



1. Description du nuage de points

Le nuage de points montre la relation entre le niveau de stress (axe des abscisses) et la durée d'écoute (axe des ordonnées). On observe une certaine dispersion des points. Visuellement, il semble qu'à des niveaux de stress plus faibles, la durée d'écoute puisse être plus variable, tandis qu'à des niveaux de stress plus élevés, la durée d'écoute semble généralement plus faible. Cependant, cette observation nécessite une confirmation statistique.

```
> plot(stress,duree_ecoute ,main="Nuage de points", xlab="Niveau de stress",
+       ylab="Durée d'écoute (en heures)")
```

2. Test de corrélation linéaire

Un test de corrélation de Pearson évalue la relation linéaire entre le niveau de stress et la durée d'écoute.

H_0 : La durée d'écoute et le niveau de stress ne sont pas corrélés linéairement.

H_1 : La durée d'écoute et le niveau de stress sont corrélés linéairement.

```
> cor.test(stress,duree_ecoute)
Pearson's product-moment correlation
data: stress and duree_ecoute
t = -5.8287, df = 76, p-value = 1.275e-07
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.6926711 -0.3803236
sample estimates:
cor
-0.5558138
```

🎵 Statistique t: -5.8287

🎵 Degrés de liberté (df): 76

🎵 p-value: 1.275e-07

🎵 Intervalle de confiance à 95 %: $[-0.6926711, -0.3803236]$

🎵 Coefficient de corrélation (cor): -0.5558138

3. Interprétation

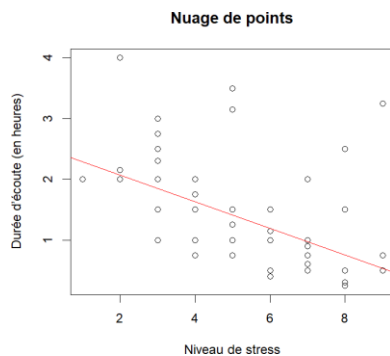
Le coefficient de corrélation linéaire ($r = -0.5558$) est négatif et modéré, indiquant une corrélation linéaire négative modérée entre le niveau de stress et la durée d'écoute. Cela suggère que lorsque le niveau de stress augmente, la durée d'écoute tend à diminuer. La p-value ($1.275e-07$) est très inférieure à 0.05, donc nous rejetons l'hypothèse nulle.

Il existe une corrélation linéaire négative significative entre le niveau de stress et la durée d'écoute ($r = -0.5558$, $p < 0.001$). Cela indique que plus le niveau de stress est élevé, moins la durée d'écoute est longue.

Régression linéaire : niveau de stress et durée d'écoute

Coefficients:

(Intercept)	stress
2.5109	-0.2198



Équation de la droite de régression :

Durée d'écoute = 2.5109 - 0.2198 * Niveau de stress

Cette équation permet de prédire la durée d'écoute en fonction du niveau de stress.

Interprétation graphique :

Le nuage de points montre la distribution des données, où chaque point représente une observation individuelle. L'axe horizontal représente le niveau de stress et l'axe vertical la durée d'écoute. La droite de régression (si tracée) traverserait ce nuage de points, indiquant la tendance générale de la relation. Dans ce cas, avec une pente négative, la droite descendrait de gauche à droite. Notez que la dispersion des points autour de cette droite peut indiquer la force de cette relation.

Interprétation des coefficients :

- ♪ **Ordonnée à l'origine (2.5109)** : C'est la valeur de la durée d'écoute estimée lorsque le niveau de stress est zéro. En d'autres termes, si une personne n'a aucun stress, on prédirait qu'elle écouterait en moyenne 2.51 heures.
- ♪ **Pente (-0.2198)** : La pente indique de combien la durée d'écoute change pour chaque augmentation d'une unité du niveau de stress. Ici, la pente est négative, ce qui signifie que pour chaque augmentation d'une unité du niveau de stress, la durée d'écoute diminue en moyenne de 0.2198 heures (environ 13 minutes). Ainsi, plus le niveau de stress est élevé, moins la personne écouterait longtemps.

Analyse bivariable entre durée d'écoute et genre préféré

1. Résumés numériques des distributions conditionnelles

Les statistiques descriptives montrent des différences dans la durée d'écoute selon les genres préférés :

```
> tapply(duree_ecoute, genre_prefere, summary)
```

- 🎵 **Rap** : moyenne = 1.76, médiane = 1.50, min = 0.50, max = 4.00, variance = 0.72
- 🎵 **Pop** : moyenne = 1.36, médiane = 1.15, min = 0.25, max = 3.25, variance = 0.55
- 🎵 **Rock** : moyenne = 1.53, médiane = 1.25, min = 0.30, max = 3.50, variance = 1.17
- 🎵 **Electro** : moyenne = 1.27, médiane = 1.00, min = 0.50, max = 2.50, variance = 0.58
- 🎵 **Classique** : moyenne = 0.57, médiane = 0.50, min = 0.50, max = 0.75, variance = 0.01
- 🎵 **Jazz** : moyenne = 1.02, médiane = 0.60, min = 0.50, max = 2.50, variance = 0.73

Ces résultats montrent que les auditeurs de Rap ont une durée d'écoute moyenne la plus élevée (1.76), tandis que ceux qui préfèrent la musique classique ont une durée significativement plus faible (0.57).

2. Rapport de corrélation (η^2)

Le rapport de corrélation (η^2) quantifie la part de variance expliquée par le genre préféré :

```
> var(duree_ecoute)
[1] 0.6978954
```

🎧 **Variance totale** : $\text{Var}(\text{durée d'écoute}) = 0.6979$

```
> varintra=sum(table(genre_prefere)*varcond)/78
> varintra
[1] 0.649472
```

🎧 **Variance intra-groupes** :

Varintra=0.6495

```
> varinter=sum(table(genre_prefere)*(mcond-mean(duree_ecoute))^2)/n
> varinter
[1] 0.08774018
```

🎧 **Variance inter-groupes** :

Varinter=0.0877

```
> varinter/var(duree_ecoute)
[1] 0.1257211
```

🎧 **Rapport η^2** :

$\eta^2 = \text{Varinter} / \text{Var} = 0.0877 / 0.6979 = 0.1257$

Interprétation : La valeur de η^2 (12.57%) indique une liaison faible entre la durée d'écoute et le genre préféré.

3. Test de comparaison ANOVA

```
> leveneTest(duree_ecoute,genre_prefere)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  5   1.424  0.226
      72
> # étape 2 : test ANOVA de comparaison des moyennes (version pour variances
> # égales grâce à l'argument var.equal=TRUE)
> oneway.test(duree_ecoute ~ genre_prefere, var.equal = TRUE)

One-way analysis of means

data: duree_ecoute and genre_prefere
F = 2.1015, num df = 5, denom df = 72, p-value = 0.07499
```

Étape 1 : Test de Levene (égalité des variances)

Hypothèses :

H_0 : Les variances des groupes sont égales.

H_1 : Les variances des groupes sont différentes.

```
> leveneTest(duree_ecoute,genre_prefere)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  5   1.424  0.226
      72
```

Résultat : p-value = 0.226 (> 0.05)

Conclusion : On ne rejette pas H_0 . Les variances sont considérées comme égales.

Étape 2 : Test ANOVA (moyennes avec variances égales)

```
> oneway.test(duree_ecoute ~ genre_prefere, var.equal = TRUE)

One-way analysis of means

data: duree_ecoute and genre_prefere
F = 2.1015, num df = 5, denom df = 72, p-value = 0.07499
```

Hypothèses :

H_0 : Les moyennes conditionnelles sont identiques pour tous les groupes.

H_1 : Au moins deux groupes ont des moyennes différentes.

Résultat : $F(5,72) = 2.10$, p-value = 0.075 (> 0.05)

Conclusion : On ne rejette pas H_0 . Il n'y a pas de différence significative entre les moyennes des groupes.

4. Comparaisons multiples

```
> # tests de comparaisons multiples pour savoir quelles moyennes diffèrent
> pairwise.t.test(duree_ecoute, genre_prefere, p.adjust.method = "none", pool.sd = TR
UE)

Pairwise comparisons using t tests with pooled SD

data: duree_ecoute and genre_prefere
```

	Rap	Pop	Rock	Electro	Classique
Pop	0.0959	-	-	-	-
Rock	0.4701	0.5662	-	-	-
Electro	0.1774	0.7879	0.5177	-	-
Classique	0.0047	0.0445	0.0332	0.1421	-
Jazz	0.0734	0.3800	0.2524	0.5963	0.3809

Bien que l'ANOVA ne soit pas significative globalement ($p > 0.05$), les tests de comparaisons multiples permettent d'explorer les différences entre les groupes :

🎵 **Rap vs Classique** : p-value < 0.005 (significatif)

♪ **Pop vs Classique** : p-value < 0.05 (significatif)

♪ **Rock vs Classique** : p-value < 0.05 (significatif)

Les autres comparaisons ne sont pas significatives.

```
> pairwise.t.test(duree_ecoute, genre_prefere, p.adjust.method = "none", pool.sd = TRUE)

Pairwise comparisons using t tests with pooled SD

data:  duree_ecoute and genre_prefere

   1      2      3      4      5
2 0.0959 -      -      -      -
3 0.4701 0.5662 -      -      -
4 0.1774 0.7879 0.5177 -      -
5 0.0047 0.0445 0.0332 0.1421 -
6 0.0734 0.3800 0.2524 0.5963 0.3809

P value adjustment method: none
```

Bien que l'ANOVA globale ne montre pas de différence significative entre les genres préférés en termes de durée d'écoute ($p > 0.05$), certaines comparaisons individuelles révèlent que les amateurs de musique classique ont une durée significativement plus faible par rapport aux amateurs de Rap, Pop et Rock. La force globale du lien reste faible ($\eta^2 < 0.2$).

Analyse bivariable entre durée d'écoute et utilisation de la musique

1. Résumés numériques des distributions conditionnelles

```
> tapply(duree_ecoute, utilisation_musique, summary)
$`1`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.400  1.000   1.500   1.374  2.000   2.500

$`2`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.300  1.000   1.500   1.506  2.000   3.150

$`3`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.250  0.500   0.500   1.309  2.000   4.000

$`4`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.500  1.000   1.000   1.325  1.413   3.500
```

Les statistiques descriptives par contexte d'utilisation montrent des moyennes similaires :

- 🎧 Pour faire du sport : moyenne = 1.37, médiane = 1.50, min = 0.40, max = 2.50, variance = 0.37
- 🎧 Pour me concentrer/travailler : moyenne = 1.51, médiane = 1.50, min = 0.30, max = 3.15, variance = 0.53
- 🎧 Pour me détendre : moyenne = 1.31, médiane = 0.50, min = 0.25, max = 4.00, variance = 1.41
- 🎧 Pendant les trajets : moyenne = 1.33, médiane = 1.00, min = 0.50, max = 3.50, variance = 0.78

Les durées moyennes oscillent entre 1.31 h et 1.51 h, avec une dispersion hétérogène (variances de 0.37 à 1.41).

2. Rapport de corrélation (η^2)

Calcul détaillé du rapport de corrélation :

- ↗ Variance totale : $\text{Var}=0.6979$
- ↗ Variance intra-groupes : $\text{Varintra}=0.7218$
- ↗ Variance inter-groupes : $\text{Varinter}=0.0067$
- ↗ Rapport η^2 : $\eta^2=\text{Varinter}/\text{Var}=0.0067/0.6979=0.0096$

Interprétation : La valeur de $\eta^2 = 0.0096$ indique une liaison quasi-nulle (0.96% de variance expliquée) entre le contexte d'utilisation et la durée d'écoute.

3. Test de comparaison ANOVA

```
> leveneTest(duree_ecoute, utilisation_musique)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  0.8108 0.4919
      74

> # étape 2 : test ANOVA de comparaison des moyennes (version pour variances
> # égales grâce à l'argument var.equal=TRUE)
> oneway.test(duree_ecoute ~ utilisation_musique, var.equal = TRUE)

One-way analysis of means

data:  duree_ecoute and utilisation_musique
F = 0.24256, num df = 3, denom df = 74, p-value = 0.8664
```

Étape 1 : Test de Levene (égalité des variances)

Hypothèses :

H_0 : Les variances sont égales entre les groupes.

H_1 : Les variances diffèrent entre au moins deux groupes.

Résultat : $p\text{-value} = 0.4919 (> 0.05)$

Conclusion : Les variances sont considérées comme égales.

Étape 2 : Test ANOVA (moyennes avec variances égales)

Hypothèses :

H_0 : Les moyennes sont identiques pour tous les groupes.

H_1 : Au moins deux groupes ont des moyennes différentes.

Résultat : $F(3,74) = 0.24256$, $p\text{-value} = 0.8664 (> 0.05)$

Conclusion : Aucune différence significative entre les moyennes des groupes.

Aucune relation significative n'est détectée entre le contexte d'utilisation et la durée d'écoute :

- 🔊 Les moyennes sont statistiquement identiques ($p = 0.8664$)
- 🔊 La force de la liaison est négligeable ($\eta^2 < 0.01$)
- 🔊 La dispersion élevée pour "Pour me détendre" ($\text{Var} = 1.41$) suggère des comportements très variables dans ce contexte.

Analyse bivariée entre durée d'écoute et sexe

1. Résumés numériques des distributions conditionnelles

Les statistiques descriptives montrent des différences dans la durée d'écoute selon le sexe :

- 🔊 Hommes : moyenne = 1.53, médiane = 1.50, min = 0.25, max = 4.00, variance = 0.73
- 🔊 Femmes : moyenne = 1.23, médiane = 1.00, min = 0.50, max = 3.50, variance = 0.63

Les hommes ont une durée moyenne légèrement plus élevée que les femmes, avec une dispersion similaire.

2. Rapport de corrélation (η^2)

Le rapport de corrélation mesure la part de variance expliquée par la variable qualitative sexe. Voici les étapes du calcul :

- 🔊 Variance totale : $\text{Var}=0.6979$
- 🔊 Variance intra-groupes : $\text{Varintra}=0.6844$
- 🔊 Variance inter-groupes : $\text{Varinter}=0.0219$
- 🔊 Rapport η^2 : $\eta^2=\text{Varinter}/\text{Var}=0.6979/0.0219=0.0314$

Interprétation : La valeur de $\eta^2 = 0.0314$ indique une liaison très faible entre le sexe et la durée d'écoute (seulement 3% de la variance totale expliquée).

3. Test de comparaison ANOVA

Étape 1 : Test de Levene (égalité des variances)

```
> leveneTest(duree_ecoute,sexe)
Levene's Test for Homogeneity of Variance (center = median)
  Df F value Pr(>F)
group 1  0.7019 0.4048
    76
```

Hypothèses :

H_0 : Les variances des groupes sont égales.

H_1 : Les variances diffèrent entre au moins deux groupes.

Résultat : p-value = 0.4048 (> 0.05)

Conclusion : On ne rejette pas H_0 . Les variances sont considérées comme égales.

Étape 2 : Test ANOVA (moyennes avec variances égales)

```
> oneway.test(duree_ecoute~sexe,var.equal=TRUE)

One-way analysis of means

data: duree_ecoute and sexe
F = 2.4968, num df = 1, denom df = 76, p-value = 0.1182
```

Hypothèses :

H_0 : Les moyennes des groupes sont identiques.

H_1 : Au moins deux groupes ont des moyennes différentes.

Résultat : $F(1,76) = 2.50$, $p\text{-value} = 0.1182 (> 0.05)$

Conclusion : On ne rejette pas H_0 . Il n'y a pas de différence significative entre les moyennes des groupes.

4. Comparaisons multiples

Non nécessaires, car l'ANOVA globale n'est pas significative.

La différence observée dans les moyennes de durée d'écoute entre hommes (1.53 h) et femmes (1.23 h) n'est pas statistiquement significative ($p > 0.05$). De plus, la liaison entre le sexe et la durée d'écoute est très faible ($\eta^2 < 0.05$). Ces résultats suggèrent que le sexe n'est pas un facteur déterminant pour expliquer les variations dans la durée d'écoute.

Analyse bivariée entre durée d'écoute et support d'écoute

1. Résumés numériques des distributions conditionnelles

Les statistiques descriptives montrent des différences dans la durée d'écoute selon le support utilisé :

```
> tapply(duree_ecoute, support_ecoute, summary)
```

- ♪ **Spotify/Deezer** : moyenne = 1.49, médiane = 1.50, min = 0.25, max = 4.00, variance = 0.72
- ♪ **Vinyle/CD** : moyenne = 0.88, médiane = 0.88, min = 0.75, max = 1.00, variance = 0.03
- ♪ **YouTube** : moyenne = 0.99, médiane = 0.50, min = 0.50, max = 2.50, variance = 0.47

Les utilisateurs de Spotify/Deezer ont une durée moyenne plus élevée (1.49 h), tandis que les utilisateurs de Vinyle/CD ont une durée nettement plus faible (0.88 h) avec une faible dispersion.

2. Rapport de corrélation (η^2)

Le rapport de corrélation mesure la part de variance expliquée par le support d'écoute :

- ♪ **Variance totale** : $\text{Var}=0.6979$
- ♪ **Variance intra-groupes** : $\text{Varintra}=0.6626$
- ♪ **Variance inter-groupes** : $\text{Varinter}=0.0419$
- ♪ **Rapport η^2** : $\eta^2=\text{Varinter}/\text{Var}=0.6979/0.0419=0.0735$

La valeur de $\eta^2 = 0.0735$ indique une liaison faible entre le support d'écoute et la durée d'écoute.

3. Test de comparaison ANOVA

Étape 1 : Test de Levene (égalité des variances)

```
> leveneTest(duree_ecoute, support_ecoute)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  1.6622 0.1966
      75
```

Hypothèses :

- 🔊 H_0 : Les variances des groupes sont égales.
- 🔊 H_1 : Les variances diffèrent entre au moins deux groupes.

Résultat : p-value = 0.1966 (> 0.05)

Conclusion : On ne rejette pas H_0 . Les variances sont considérées comme égales.

Étape 2 : Test ANOVA (moyennes avec variances égales)

```
> oneway.test(duree_ecoute ~ support_ecoute, var.equal = TRUE)
One-way analysis of means
data:  duree_ecoute and support_ecoute
F = 2.4313, num df = 2, denom df = 75, p-value = 0.09483
```

Hypothèses :

- 🔊 H_0 : Les moyennes des groupes sont identiques.
- 🔊 H_1 : Au moins deux groupes ont des moyennes différentes.

Résultat : $F(2,75) = 2.43$, p-value = 0.0948 (> 0.05)

Conclusion : On ne rejette pas H_0 . Il n'y a pas de différence significative entre les moyennes des groupes.

4. Comparaisons multiples

```
> pairwise.t.test(duree_ecoute, support_ecoute, p.adjust.method = "none", pool.sd = TRUE)

Pairwise comparisons using t tests with pooled SD
data:  duree_ecoute and support_ecoute
  1      2
2 0.296 -
3 0.048 0.851
```

Bien que l'ANOVA globale ne soit pas significative, les tests par paires montrent :

- 🔊 Spotify/Deezer vs YouTube : p-value = 0.048 (significatif)
- 🔊 Les autres comparaisons ne sont pas significatives.

Bien que l'ANOVA globale ne révèle pas de différence significative entre les supports d'écoute ($p > 0.05$), les comparaisons multiples suggèrent que les utilisateurs de Spotify/Deezer écoutent significativement plus longtemps que ceux de YouTube. La liaison globale reste faible ($\eta^2 < 0.10$).

Analyse bivariable entre durée d'écoute et moment d'écoute

1. Résumés numériques des distributions conditionnelles

Les statistiques descriptives montrent des différences dans la durée d'écoute selon le moment d'écoute :

```
> tapply(duree_ecoute, moment_ecoute, summary)
```

🔊 **Le matin** : moyenne = 1.19, médiane = 1.00, min = 0.25, max = 3.50, variance = 0.63

🔊 **L'après-midi** : moyenne = 1.21, médiane = 1.00, min = 0.50, max = 2.50, variance = 0.42

🔊 **Le soir** : moyenne = 1.57, médiane = 1.50, min = 0.50, max = 4.00, variance = 0.81

Les auditeurs du soir ont une durée moyenne plus élevée (1.57 h) avec une dispersion plus importante par rapport aux autres moments.

2. Rapport de corrélation (η^2)

Le rapport de corrélation mesure la part de variance expliquée par le moment d'écoute :

🔊 **Variance totale** : Var=0.6979

🔊 **Variance intra-groupes** : Varintra=0.6644

🔊 **Variance inter-groupes** : Varinter=0.0334

🔊 **Rapport η^2** : $\eta^2 = \text{Varinter} / \text{Var} = 0.0334 / 0.6979 = 0.0479$

La valeur de $\eta^2 = 0.0479$ indique une liaison très faible entre le moment d'écoute et la durée d'écoute.

3. Test de comparaison ANOVA

Étape 1 : Test de Levene (égalité des variances)

```
> leveneTest(duree_ecoute, moment_ecoute)
Levene's Test for Homogeneity of Variance (center = median)
  Df F value Pr(>F)
group 2  1.0716 0.3478
    73
```

Hypothèses :

- ♪ H_0 : Les variances des groupes sont égales.
- ♪ H_1 : Les variances diffèrent entre au moins deux groupes.

Résultat : p-value = 0.3478 (> 0.05)

Conclusion : On ne rejette pas H_0 . Les variances sont considérées comme égales.

Étape 2 : Test ANOVA (moyennes avec variances égales)

```
> oneway.test(duree_ecoute ~ moment_ecoute, var.equal = TRUE)

One-way analysis of means

data:  duree_ecoute and moment_ecoute
F = 1.902, num df = 2, denom df = 73, p-value = 0.1566
```

Hypothèses :

- 🎧 H_0 : Les moyennes des groupes sont identiques
- 🎧 H_1 : Au moins deux groupes ont des moyennes différentes.

Résultat : $F(2,73) = 1.902$, p-value = 0.1566 (> 0.05)

Conclusion : On ne rejette pas H_0 . Il n'y a pas de différence significative entre les moyennes des groupes.

4. Comparaisons multiples

```
> pairwise.t.test(duree_ecoute, moment_ecoute, p.adjust.method = "none", pool.sd = TRUE)

Pairwise comparisons using t tests with pooled SD

data:  duree_ecoute and moment_ecoute

   1      2
2 0.93 -
3 0.10 0.15

P value adjustment method: none
```

Bien que l'ANOVA globale ne soit pas significative, les tests par paires montrent :

- ♪ **Le matin vs Le soir** : p-value = 0.10 (non significatif)
- ♪ **L'après-midi vs Le soir** : p-value = 0.15 (non significatif)

Aucune comparaison n'est significative.

La différence observée dans les moyennes de durée d'écoute entre les moments d'écoute (matin, après-midi, soir) n'est pas statistiquement significative ($p > 0.05$). De plus, la liaison entre le moment d'écoute et la durée est très faible ($\eta^2 < 0.05$).

Analyse bivariée entre fréquence musique et utilisation de la musique (Quali/quali)

1. Tableau de contingence

```
> table_cont
      utilisation_musique
frequence_ecoute 1  2  3  4
1      13 14  5 11
2       6  9  7  1
3       1  3  5  2
```

Le tableau de contingence présente les effectifs observés pour chaque combinaison des deux variables.

- ♪ Les auditeurs quotidiens utilisent principalement la musique pour se concentrer/travailler (14) et faire du sport (13).
- ♪ Les auditeurs presque quotidiens privilégient la détente (7) et le travail/concentration (9).
- ♪ Les auditeurs quelques fois par semaine écoutent majoritairement la musique pour se détendre (5).

Les totaux marginaux montrent que l'utilisation principale est se concentrer/travailler (26), suivie de faire du sport (20).

2. Distributions conditionnelles

Profil-lignes (sachant la fréquence d'écoute)

```
> round(prop.table(table_cont,1),digits=3)
      utilisation_musique
frequence_ecoute 1  2  3  4
1      0.302 0.326 0.116 0.256
2      0.261 0.391 0.304 0.043
3      0.091 0.273 0.455 0.182
```

Les proportions indiquent comment l'utilisation musicale est répartie pour chaque fréquence d'écoute :

Auditeurs quotidiens :

- ♪ 30.2% utilisent la musique pour faire du sport.
- ♪ 32.6% pour se concentrer/travailler.
- ♪ 25.6% pendant les trajets.

Auditeurs presque quotidiens :

- ♪ Utilisation principale : se concentrer/travailler (39.1%).
- ♪ Deuxième usage : se détendre (30.4%).

Auditeurs quelques fois par semaine :

- ♪ Usage dominant : se détendre (45.5%).

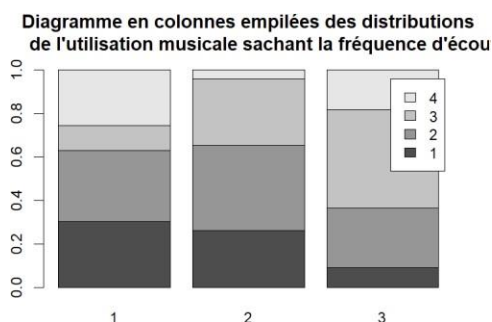
Profil-colonnes (sachant l'utilisation musicale)

```
> round(prop.table(table_cont,2),digits=3)
      utilisation_musique
frequence_ecoute  1      2      3      4
1 0.650 0.538 0.294 0.786
2 0.300 0.346 0.412 0.071
3 0.050 0.115 0.294 0.143
```

Les proportions montrent comment la fréquence d'écoute est répartie selon l'utilisation musicale :

- 🎵 **Faire du sport** : 65% des auditeurs sont des quotidiens.
- 🎵 **Se concentrer/travailler** : 53.8% sont des quotidiens, mais 34.6% écoutent presque tous les jours.
- 🎵 **Se détendre** : Répartition plus équilibrée entre :
 - 🎧 29.4% quotidiens
 - 🎧 41.2% presque quotidiens
 - 🎧 29.4% quelques fois par semaine.

3. Diagramme en colonnes empilées des Profils-Lignes



1. **Tous les jours** : Les auditeurs quotidiens répartissent leurs utilisations de manière équilibrée entre faire du sport (30.2%), se concentrer/travailler (32.6%), et pendant les trajets (25.6%), avec une moindre utilisation pour se détendre (11.6%).
2. **Presque tous les jours** : Cette catégorie privilégie se concentrer/travailler (39.1%) et se détendre (30.4%), tandis que l'utilisation pour pendant les trajets est très faible (4.3%).
3. **Quelques fois par semaine** : Les auditeurs occasionnels utilisent principalement la musique pour se détendre (45.5%), avec une part notable pour se concentrer/travailler (27.3%), et des utilisations marginales pour le sport et les trajets.

Ces observations montrent que la fréquence d'écoute influence directement les préférences d'utilisation musicale, avec des auditeurs réguliers ayant des usages variés et des auditeurs occasionnels se concentrant sur des activités relaxantes comme la détente.

4. Mosaic Plot en Profil-Colonnes



1. **Dominance des auditeurs quotidiens** : Les auditeurs *tous les jours* dominant dans les catégories "faire du sport" (65%) et "pendant les trajets" (78.6%), indiquant une forte association entre ces utilisations et une écoute régulière.
2. **Diversité dans "se détendre"** : L'utilisation pour "se détendre" est la plus équilibrée, avec une répartition notable entre *tous les jours* (29.4%), *presque tous les jours* (41.2%), et *quelques fois par semaine* (29.4%).
3. **Équilibre pour "se concentrer/travailler"** : Bien que dominée par les auditeurs quotidiens (53.8%), cette catégorie inclut une proportion significative d'auditeurs presque quotidiens (34.6%), reflétant une utilisation moins exclusivement liée à une écoute quotidienne.

5. Test du χ^2

Hypothèses

- ♫ H_0 : La fréquence d'écoute et l'utilisation musicale sont indépendantes.
- ♫ H_1 : Il existe une dépendance entre la fréquence d'écoute et l'utilisation musicale.

Résultat

Pearson's Chi-squared test

data: frequence_ecoute and utilisation_musique
 X-squared = 11.152, df = 6, p-value = 0.08378

Le test donne :

- ♫ **p-value = 0.08378 (< 0.1).**

La p-value étant significative à 10%, on rejette H_0 .

Il y a un lien significatif entre la fréquence d'écoute et l'utilisation musicale.

6. Tableau des contributions au Khi²

```
> khi2$residuals^2 # contributions au Khi2 c_ij
      utilisation_musique
frequence_ecoute 1      2      3      4
1 0.3002265177 0.0185860651 2.1268854265 1.2949260042
2 0.0001129305 0.1959996525 0.7275384462 2.4209486166
3 1.2071428571 0.1373626374 2.7226890756 0.0000000000
> khi2$expected # effectifs théoriques t_ij
      utilisation_musique
frequence_ecoute 1      2      3      4
1 11.168831 14.519481 9.493506 7.818182
2 5.974026 7.766234 5.077922 4.181818
3 2.857143 3.714286 2.428571 2.000000
> khi2$observed # effectifs observés n_ij (=tableau de contingence)
      utilisation_musique
frequence_ecoute 1 2 3 4
1 13 14 5 11
2 6 9 7 1
3 1 3 5 2
```

Plus grande sur-représentation : Fréquence "Quelques fois/semaine" / Utilisation "Pour se détendre"

Contribution au Khi² : 2.72

Effectif observé : 5

Effectif attendu : 2.43

Les auditeurs qui écoutent de la musique quelques fois par semaine utilisent ce temps principalement pour se détendre donc plus souvent que ce qui serait attendu en supposant l'indépendance entre fréquence d'écoute et utilisation de la musique. Cela suggère une forte association entre une écoute moins fréquente et une utilisation relaxante de la musique. Les auditeurs privilégient des moments spécifiques de détente musicale.

QUESTIONNAIRE :

https://docs.google.com/forms/d/e/1FAIpQLScAQ_i1dYBTcSPISovsj7GZ7nHEpT-P4QtU4n4XXRjlt8NXlg/viewform?pli=1&pli=1