



17 November 2018

Alexandre Allani
Axel Durand
Adrien Clotteau
Mathieu Roy
Matthieu Bachelot

Projet Statistique groupe 3

La culture cinématographique



IMT Atlantique

Bretagne-Pays de la Loire
École Mines-Télécom

Contents

1	Contexte	4
1.1	La création du questionnaire	4
1.2	Le choix d'échantillonnage	5
1.3	Étude de l'échantillon	5
2	Analyse statistiques de l'échantillon: les hypothèses et leur vérification	6
2.1	Hypothèse 1	6
2.1.1	Visualisation	6
2.1.2	Application du test	7
2.2	Hypothèses 2.1 et 2.2	7
2.2.1	Visualisation	7
2.2.2	Application des test	8
2.2.2.1	Test hypothèse 2.1:	8
2.2.2.2	Test hypothèse 2.2:	8
2.2.3	Conclusion des tests	8
2.3	Hypothèses 3.1 et 3.2	9
2.3.1	Hypothèse 3.1 :	9
2.3.1.1	Application des test	9
2.3.2	Hypothèse 3.2:	10
2.3.2.1	Application des test	10
2.3.3	Conclusion des tests	10
2.4	Hypothèses 4.1 et 4.2	10
2.4.1	Hypothèse 4.1	10
2.4.1.1	Application du test	11
2.4.2	Hypothèse 4.2	11
2.4.2.1	Application des test	11
2.4.3	Conclusion	11
2.5	Hypothèse 5	11
2.5.1	Visualisation	12
2.5.2	Application des test	12
2.6	Hypothèse 6	12
2.6.1	Visualisation	13
2.6.2	Application des test	13
2.7	Hypothèse 7	14
2.7.1	Visualisation	14
2.7.2	Application des test	15
2.7.2.1	H7.1: Les personnes n'ayant pas vu Star Wars reconnaissent aussi bien que celles qui l'ont vu.	15
2.7.2.2	H7.2: Les personnes n'ayant pas vu le Roi Lion reconnaissent aussi bien que celles qui l'ont vu.	15
2.7.2.3	H7.3: Les personnes n'ayant pas vu Terminator reconnaissent aussi bien que celles qui l'ont vu.	16
2.7.2.4	H7.4: Les personnes n'ayant pas vu Titanic reconnaissent aussi bien que celles qui l'ont vu.	16
2.7.2.5	H7.5: Les personnes n'ayant pas vu le Seigneur des anneaux reconnaissent aussi bien que celles qui l'ont vu.	16
2.7.3	Conclusion	16

- 2.8 Hypothèse 8 16
 - 2.8.1 Visualisation 17
 - 2.8.2 Application des test 17

“Les cons ça ose tout, c’est même à ça qu’on les reconnaît.”

Les tontons flingueurs
Michel Audiard

1 Contexte

Le choix de thème de notre sondage s’est porté sur le cinéma pour deux raisons majeures :

- Nous cherchions un thème qui intéresse les sondés pour recueillir les réponses de la part d’un maximum de personnes et ainsi pouvoir faire des statistiques sans contraintes de taille d’échantillon.
- Nous cherchions un thème qui nous parlait à tous, qui aurait été amusant à réaliser et dont les résultats nous auraient intéressés.

Le premier point évoqué a été très important dans l’élaboration de notre questionnaire. Nous avons voulu le réaliser de manière originale, attractive et agréable à compléter afin de capter et de garder jusqu’au bout autant de sondés que possible.

1.1 La création du questionnaire

Le questionnaire a été réalisé en 3 parties.

1. Le recueil des informations personnelles du sondé :
 - (a) Âge(Nombre)
 - (b) Sexe (homme/femme)
 - (c) Pays d’origine (Choix dans la liste des pays)
 - (d) Avez-vous étudié à IMT Atlantique (anciennement Telecom Bretagne) ? (Oui/Non)
 - (e) Pensez-vous reconnaître plus un film à ses répliques ou à sa musique ? (Musique/Répliques)
2. Un blind test audio et visuel sur 5 films estimés comme majeurs dans l’univers cinématographique (Star Wars, le Seigneur des Anneaux, Titanic, le Roi Lion et Terminator). Par son côté ludique, celui-ci augmente l’intérêt des sondés et les motive à aller jusqu’au bout du questionnaire. Ceux-ci doivent reconnaître une réplique et une musique tirées des 5 films précédemment cités. Les dix tests (5 films et 5 musiques) sont réalisés dans un ordre que nous avons choisi pour éviter que la musique et la réplique d’un même film ne se suivent dans le sondage. Les extraits musicaux durent 20 secondes et peuvent être répétés. Pour chaque réplique et musique, les questions suivantes sont posées.
 - (a) Connaissez-vous la référence ? (Oui/Non)
 - (b) (Si oui) Dans quel film l’avez vous entendu ? (Nom à marquer)
3. Un sondage après révélation des films. Les cinq films sont présentés avec leurs affiches. En dessous de chaque affiche, le sondé doit répondre aux questions suivantes :
 - (a) Combien de fois avez-vous vu ce film ? (0 fois/1 fois/2-5 fois/Plus de 5 fois)
 - (b) (Si plus d’une fois) À quel âge avez-vous vu ce film pour la première fois ? (0-10 ans/11-20 ans/Plus de 20 ans)
4. Un notation de la part du sondé des caractéristiques qu’il juge essentielles pour qu’un film soit mémorable. Six caractéristiques sont données :
 - (a) Répliques
 - (b) Musique
 - (c) Qualité de l’histoire
 - (d) Revoir un film plusieurs fois sans se lasser
 - (e) L’avoir vu enfant

(f) Porteur d'émotion

Pour chaque caractéristique, le sondé devra choisir entre : D'accord/Plutôt d'accord/Neutre/Plutôt pas d'accord/Pas d'accord.

Les 5 films que nous avons sélectionnés correspondent à des films de différents genres destinés à différents publics et sortis lors de différentes décennies (Star Wars : 1977, le Seigneur des Anneaux : 2001, Titanic : 1997, le Roi Lion : 1994 et Terminator : 1984). Cependant, ils correspondent à des films cultes selon les critères de notre génération et peuvent ne pas être considérés comme tels par les autres.

1.2 Le choix d'échantillonnage

N'ayant pas un large choix de sondeurs (étudiants d'IMT Atlantique, familles etc ...), nous avons fait le choix de ne pas échantillonner nos sondés pour avoir le plus de réponses possibles. De plus, notre thème est censé devoir toucher tout type de population. Il a été donc important, dans notre sondage, de demander les caractéristiques des sondés (âge, nationalité, étudiant d'IMT).

1.3 Étude de l'échantillon

Nous avons obtenu 388 réponses au sondage. 86 d'entre elles n'étaient pas complètes. Deux autres données étaient obsolètes, ce qui ramène le nombre de réponses traitées à 300. Parmi ces données, deux individus ont donné des âges incohérents, ils ne seront donc pas étudiés pour les statistiques par rapport à l'âge.

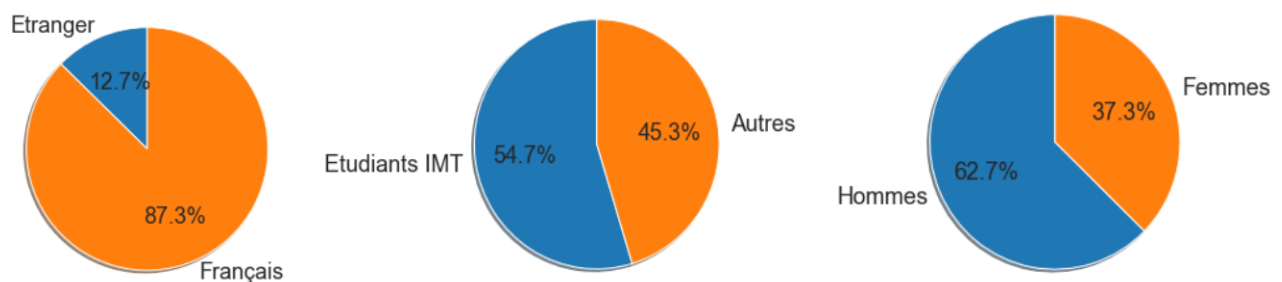


Figure 1: Caractéristiques des sondés

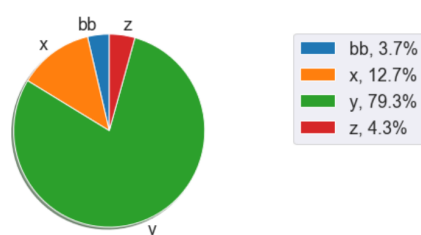


Figure 2: Générations auxquelles appartiennent les sondés

des étudiants d'IMT Atlantique et un faible nombre de baby-boomers (plus de 54 ans) et de générations Z (moins de 18 ans).

On peut remarquer une grande proportion d'étudiants IMT, de Français et d'hommes dans nos échantillons. De plus, les réponses à la question "Pensez-vous reconnaître plus un film par ses musiques ou par ses répliques ?" sont équitablement réparties (52,3% pour les répliques contre 47,7% pour les musiques).

On peut aussi remarquer une grande proportion d'individus de la génération Y (18-38 ans) qui correspond à la tranche d'âge

2 Analyse statistiques de l'échantillon: les hypothèses et leur vérification



Vous pouvez retrouver une version plus détaillée de l'analyse des résultats (avec le code) en cliquant sur le lien suivant : <https://github.com/Syndorik/ProjetStats/blob/master/Projet%20Stats.html>. Il faut télécharger le fichier afin de pouvoir visualiser les graphiques

Dans la partie qui suit, nous allons énumérer les différentes hypothèses que nous avons posées, ainsi que l'analyse statistique faite pour les valider (ou invalider).

Les lois de distribution de nos variables ne sont pas normalement distribuées et sont discrètes. Cependant, par le théorème de centrale limite, vu que notre échantillon n est grand ($n = 300$), la distribution des moyennes tend vers une loi normale. On utilise ce théorème pour toutes nos hypothèses.

2.1 Hypothèse 1

H_0 = Les sondés reconnaissent autant un film par sa musique que par sa réplique.

➤ $H_0 : \mu_{musique} = \mu_{réplique}$

➤ $H_1 : \mu_{musique} \neq \mu_{réplique}$

Le but ici est, comme l'indique l'intitulé de l'hypothèse, de comparer les taux de bonnes réponses au blind test, pour les musiques et pour les répliques.

2.1.1 Visualisation

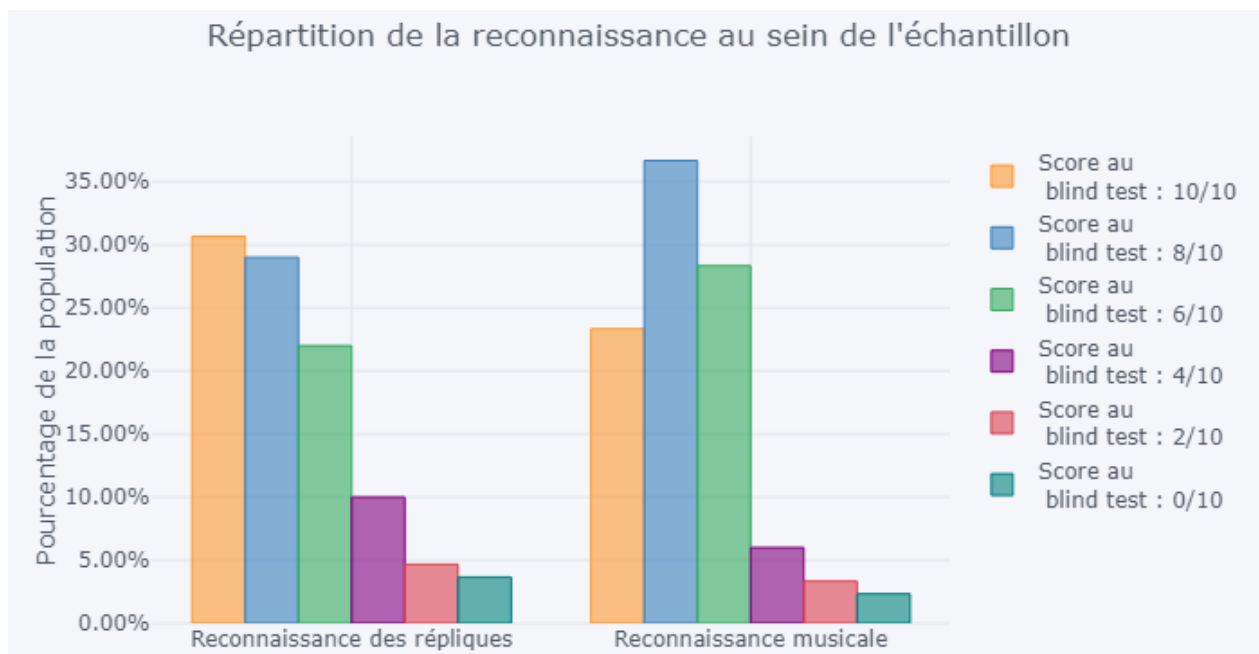


Figure 3: Pourcentage de reconnaissance des musiques et répliques

La répartition correspond à nos attentes. Nous cherchions des films que l'ensemble des sondés pouvaient connaître, mais qui n'étaient pas non plus reconnus par tout le monde. On peut remarquer qu'il y a plus de personnes ayant reconnu l'ensemble des répliques que des musiques. Cependant il y a plus de gens ayant reconnus au moins 3 musique que de gens ayant reconnu au moins 3 répliques.

2.1.2 Application du test

En regardant notre hypothèse, avec l'hypothèse nulle et l'hypothèse alternative, on remarque que l'on peut utiliser ici un test de Student. Par ailleurs les données sont appariées (la musique et la réplique sont associées pour une même personnes). De plus, les échantillons sont bien indépendants.

La moyenne de reconnaissance des répliques est de 0.73 et la moyenne de reconnaissance des musiques est de 0.72. Immédiatement, en regardant le résultat, on peut s'attendre à ce que le test confirme l'hypothèse nulle.

Après application du test, nous obtenons les valeurs suivantes : $pvalue = 0.5817$, $t = 0.5514$. Donc $pvalue > 0.05$. De plus, l'intervalle de confiance des différences de moyenne est de : $[-0.05, 0.03]$. La différence des moyennes observées appartenant à l'échantillon est en accord avec la $pvalue$ calculée



On peut conclure que les sondés reconnaissent autant un film par sa musique que par sa réplique.

2.2 Hypothèses 2.1 et 2.2

H2.1 : les hommes reconnaissent plus les films par la réplique que les femmes. De nouveau nous faisons un test de Student, avec pour hypothèses:

- $H_0 : \mu_{homme_rep} = \mu_{femme_rep}$
- $H_a : \mu_{homme_rep} \neq \mu_{femme_rep}$

H2.2 : Les femmes reconnaissent plus les films par les musiques que les hommes. Test de Student, avec pour hypothèses:

- $H_0 : \mu_{homme_mus} = \mu_{femme_mus}$
- $H_a : \mu_{homme_mus} \neq \mu_{femme_mus}$

Le but ici est de comparer la reconnaissance des films par les hommes et les femmes. Pour les deux hypothèses, nous allons appliquer le test de Student.

2.2.1 Visualisation

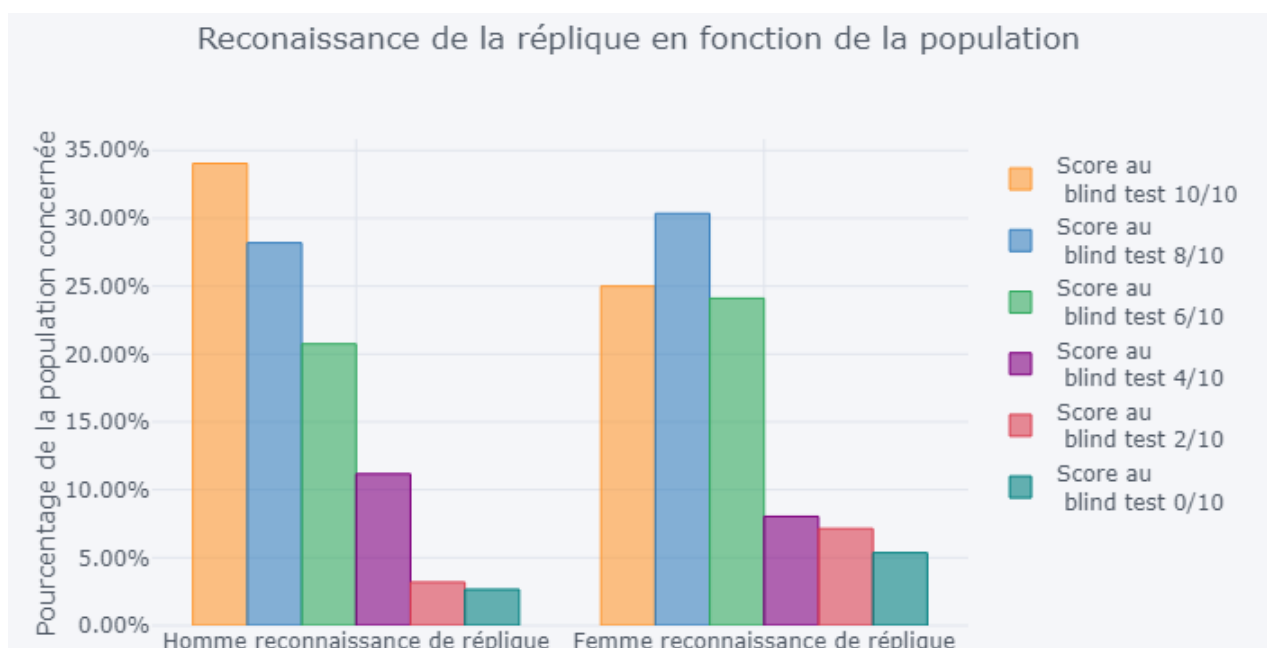


Figure 4: Pourcentage de reconnaissance des répliques pour les hommes et les femmes

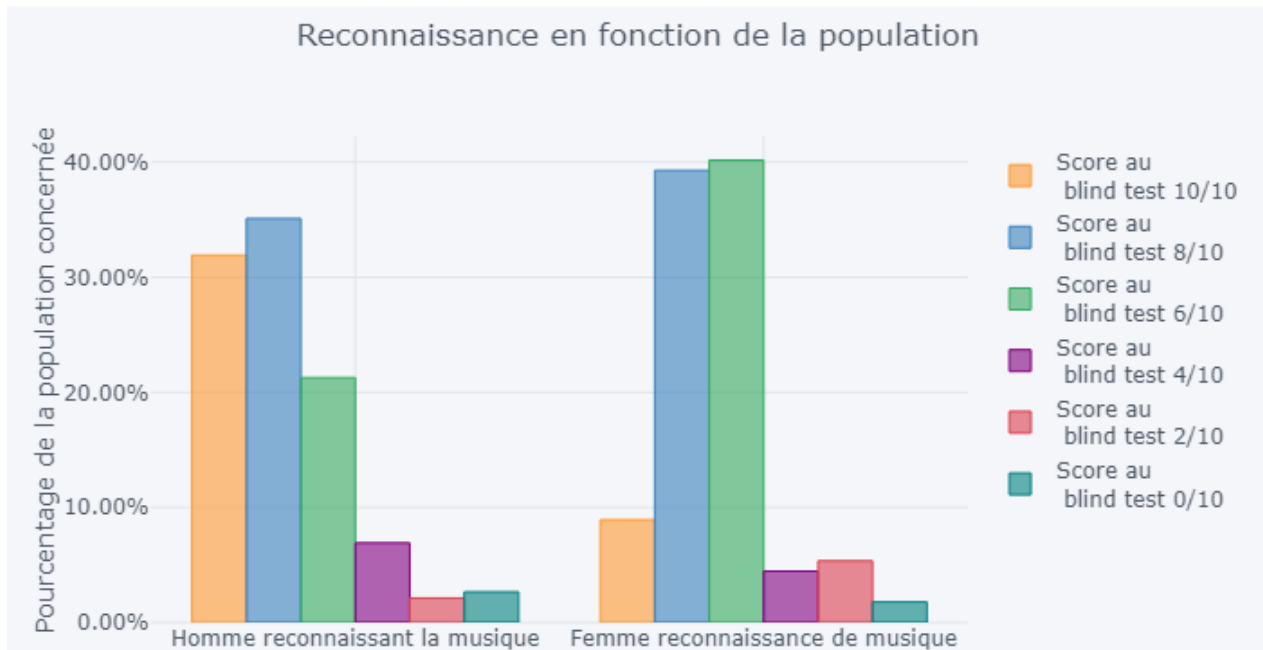


Figure 5: Pourcentage de reconnaissance des musiques pour les hommes et les femmes

La distribution entre les hommes et les femmes est assez différente. En particulier les hommes ont l'air de reconnaître mieux à la réplique contrairement aux femmes. Pareillement, les hommes ont l'air de reconnaître mieux la musique que les femmes.

2.2.2 Application des test

2.2.2.1 Test hypothèse 2.1:

Les hommes reconnaissent plus les films par la réplique que les femmes

Après application du test, nous obtenons les valeurs suivantes : $pvalue = 0.070$, $t = 1.814$. Donc $pvalue > 0.05$. De plus, l'intervalle de confiance des différences de moyenne est de : $[-0.12, 0.01]$.



On ne peut pas affirmer l'hypothèse nulle. En effet, la p-value qu'on obtient après le test de student est de 0.07. Cependant, la pvalue étant très proche de 0.05, utilisé communément en statistique pour rejeter l'hypothèse, nous pouvons en déduire qu'il y a de forte présomption contre l'hypothèse.

2.2.2.2 Test hypothèse 2.2:

Les femmes reconnaissent plus les films par la musique que les hommes. Après application du test, nous obtenons les valeurs suivantes : $pvalue = 0.001$, $t = 3.219$. Donc $pvalue < 0.05$. De plus, l'intervalle de confiance des différences de moyenne est de : $[-0.14, -0.04]$.



On peut rejeter l'hypothèse nulle via la pvalue.

2.2.3 Conclusion des tests

Les hommes reconnaissent mieux les répliques et les musiques que les femmes. Ce résultat pouvait être anticipable. En effet, comme nous l'expliquerons dans les biais du sondage un peu plus tard, dans notre

groupe nous sommes 6 hommes, et donc notre choix de film a été influencé, avec plus de chance d'être réussi par d'autres hommes.

2.3 Hypothèses 3.1 et 3.2

Nous comparons ici les femmes ayant répondu différemment à la question préliminaire "Pensez-vous reconnaître un film plus à ses répliques ou à sa musique?". L'intérêt de cette hypothèse est de voir si les femmes savent comment elles vont reconnaître les films.

Pour les deux hypothèses, nous appliquons un test de Student.

H3.1 : Les femmes pensant reconnaître les films plus par la musique reconnaissent autant la musique que celles qui pensent reconnaître les films par la réplique.

➤ $H_0 : \mu_{femme_recmus} = \mu_{femme_recrep}$

➤ $H_a : \mu_{femme_recmus} \neq \mu_{femme_recrep}$

Ici μ_{femme_recmus} représente la moyenne des réponses (sur la musique) des femmes pensant reconnaître la musique. μ_{femme_recrep} représente la moyenne des réponses (sur la musique) des femmes pensant reconnaître la réplique.

2.3.1 Hypothèse 3.1 :

2.3.1.1 Application des test

La moyenne de femme pensant reconnaître la musique et qui reconnaissent la musique est de 0.68. La moyenne de femme pensant reconnaître la réplique et qui reconnaissent la musique est de 0.67. Après application du test, nous obtenons les valeurs suivantes : $pvalue = 0.7452$, $t = 0.3257$. Donc $pvalue > 0.05$. De plus, l'intervalle de confiance des différences de moyenne est de : $[-0.09, 0.06]$. La différence des moyennes observées appartenant à l'échantillon est en accord avec la pvalue calculée.



Avec cette valeur de pvalue, nous pouvons affirmer l'hypothèse nulle.



Dans notre questionnaire nous n'avons pas mis de case "Je reconnais un film par la musique et par les répliques". Ainsi une personne ayant tout reconnu en film et en réplique ferait plutôt parti de cette population. Cela biaise un peu le résultat précédent.



On applique de nouveau un test de Student mais cette fois-ci sans prendre en compte celles qui ont fait un 100% au blind test afin d'éliminer ce biais

La moyenne de femme pensant reconnaître la musique et qui reconnaissent la musique est de 0.66. La moyenne de femme pensant reconnaître la réplique et reconnaissant la musique est de 0.65. Après application du test, nous obtenons les valeurs suivantes : $pvalue = 0.7657$, $t = 0.2986$. Donc $pvalue > 0.05$. De plus, l'intervalle de confiance des différences de moyenne est de : $[-0.09, 0.06]$. La différence des moyennes observées appartenant à l'échantillon est en accord avec la pvalue calculée.



Avec cette valeur de pvalue, nous devons toujours confirmer l'hypothèse nulle (le biais est toujours présent). Ainsi, avoir choisi "musique" ou "réplique" dans la question préliminaire n'implique pas une différenciation de bonnes réponses sur la musique, pour les femmes.

2.3.2 Hypothèse 3.2:

H3.2 : Les femmes pensant reconnaître les films plus par la réplique reconnaissent autant la réplique que celles qui pensent reconnaître les films par la musique.

➤ $H_0 : \mu_{femme_recmus} = \mu_{femme_recrep}$

➤ $H_a : \mu_{femme_recmus} \neq \mu_{femme_recrep}$

Ici μ_{femme_recmus} représente la moyenne des réponses (sur la réplique) des femmes pensant reconnaître la musique. μ_{femme_recrep} représente la moyenne des réponses (sur la réplique) des femmes pensant reconnaître la réplique.

2.3.2.1 Application des test

On choisit immédiatement de ne pas prendre en compte les femmes ayant répondu bon à 100% au blind test.

La moyenne de femme pensant reconnaître la musique et reconnaissant les répliques est de 0.63. La moyenne de femme pensant reconnaître la réplique et reconnaissant les répliques est de 0.71.

Après application du test, nous obtenons les valeurs suivantes : $pvalue = 0.1456$, $t = -1.466$. Donc $pvalue > 0.05$. De plus, l'intervalle de confiance des différences de moyenne est de : $[-0.03, 0.19]$.



Après application du test, nous devons accepter l'hypothèse nulle. La pvalue, bien que plus faible qu'au test précédent, reste supérieur à 0.05.

2.3.3 Conclusion des tests

Avec ces tests, nous pouvons affirmer que la réponse à la question préliminaire est très peu pertinente pour les femmes, dans ce sondage. Le fait que nos films soient peu nombreux et plutôt "simple" à reconnaître empêche un clivage marquant avec la question préliminaire.

2.4 Hypothèses 4.1 et 4.2

Nous comparons ici les hommes ayant répondu différemment à la question préliminaire "Pensez-vous reconnaître un film plus à ses répliques ou à sa musique?". L'intérêt de cette hypothèse est de voir si les hommes savent comment ils vont reconnaître les films. Pour les deux hypothèses, nous appliquons un test de Student.



En continuation avec les tests sur les femmes, nous choisissons d'exclure les hommes ayant eu 100% au blind test.

2.4.1 Hypothèse 4.1

H4.1 : Les hommes pensant reconnaître les films plus par la musique reconnaissent autant la musique que ceux qui pensent reconnaître les films par la réplique.

➤ $H_0 : \mu_{homme_recmus} = \mu_{homme_recrep}$

➤ $H_a : \mu_{homme_recmus} \neq \mu_{homme_recrep}$

Ici μ_{homme_recmus} représente la moyenne des réponses (sur la musique) des hommes pensant reconnaître la musique. μ_{homme_recrep} représente la moyenne des réponses (sur la musique) des hommes pensant reconnaître la réplique.

2.4.1.1 Application du test

La moyenne des hommes pensant reconnaître la musique et reconnaissant la musique est de 0.72. La moyenne des hommes pensant reconnaître la réplique et reconnaissant la musique est de 0.67. Après application du test, nous obtenons les valeurs suivantes : $pvalue = 0.2054$, $t = 1.271$. Donc $pvalue < 0.05$. De plus, l'intervalle de confiance des différences de moyenne est de : $[-0.12, 0.03]$.



Avec cette valeur de $pvalue$, nous pouvons affirmer l'hypothèse nulle.

2.4.2 Hypothèse 4.2

H4.2 : Les hommes pensant reconnaître les films plus par la réplique reconnaissent autant la réplique que ceux qui pensent reconnaître les films par la musique.

➤ $H_0 : \mu_{homme_recmus} = \mu_{homme_recrep}$

➤ $H_a : \mu_{homme_recmus} = \mu_{homme_recrep}$

Ici μ_{homme_recmus} représente la moyenne des réponses (sur la réplique) des hommes pensant reconnaître la musique. μ_{homme_recrep} représente la moyenne des réponses (sur la réplique) des hommes pensant reconnaître la réplique.

2.4.2.1 Application des test La moyenne des hommes pensant reconnaître la musique et reconnaissant la réplique est de 0.68. La moyenne des hommes pensant reconnaître la réplique et reconnaissant la réplique est de 0.67.

Après application du test, nous obtenons les valeurs suivantes : $pvalue = 0.8612$, $t = 0.1751$. Donc $pvalue > 0.05$. De plus, l'intervalle de confiance des différences de moyenne est de : $[-0.09, 0.07]$.



Avec cette valeur de $pvalue$, nous pouvons affirmer l'hypothèse nulle.

2.4.3 Conclusion

Avec ces tests, nous pouvons affirmer que la réponse à la question préliminaire est très peu pertinente pour les hommes aussi, dans ce sondage.

2.5 Hypothèse 5

H5 : Les français reconnaissent les films proposés autant que les autres nationalités.

➤ $H_0 : \mu_{nationalit} = \mu_{français}$

➤ $H_a : \mu_{nationalit} \neq \mu_{français}$

Nous appliquons de nouveau le test de Student.

2.5.1 Visualisation

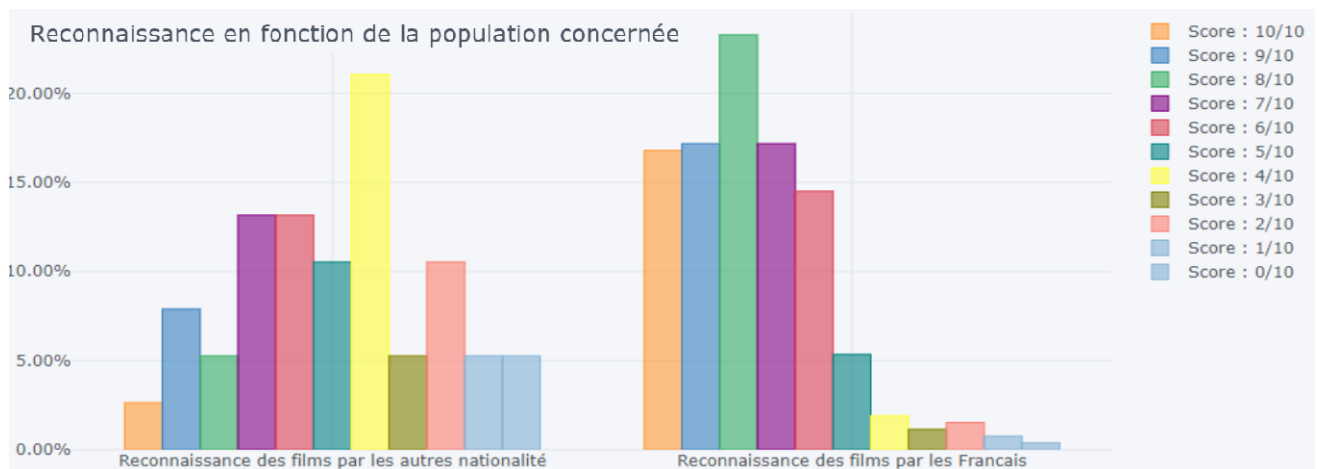


Figure 6: Score au blind test en fonction de la nationalité

Ce graphe nous permet immédiatement de prévoir le résultat flagrant du test, c'est à dire la validation de l'hypothèse alternative.

2.5.2 Application des test

La moyenne de reconnaissance de film par les français est de 0.76 et la moyenne de reconnaissance de film par les autres nationalités est de 0.49.

Après application du test, nous obtenons les valeurs suivantes : $pvalue = 2.3920e^{-13}$, $t = 7.674$. Donc $pvalue < 0.05$. De plus, l'intervalle de confiance des différences de moyenne est de : $[-0.36, -0.18]$.



Après application du test, on peut rejeter l'hypothèse nulle. En effet la pvalue est presque nulle donc inférieure à 0.05. Ainsi, nous avons la confirmation que les français ont mieux répondu au blind test que les étrangers. Nous avons conscience de ce biais à la création du sondage et nous voulions voir s'il était confirmé. Cela est bien le cas

2.6 Hypothèse 6

H5.5 : Parmi les différentes générations (baby-boomers / X / Y / Z), la génération Y (née entre 1980 et 2000) a une meilleure connaissance des films qui sont proposés.

- $H_0 : \mu_{babyboomer} = \mu_X = \mu_Y = \mu_Z$
- H_a : Il y a une différence parmi les moyennes

On va utiliser ANOVA pour confirmer ou infirmer les hypothèses. De même que pour Student, bien que nos variables prennent des valeurs discrètes et leur distribution n'est pas normale, on peut appliquer l'ANOVA grâce à la taille de l'échantillon.

2.6.1 Visualisation

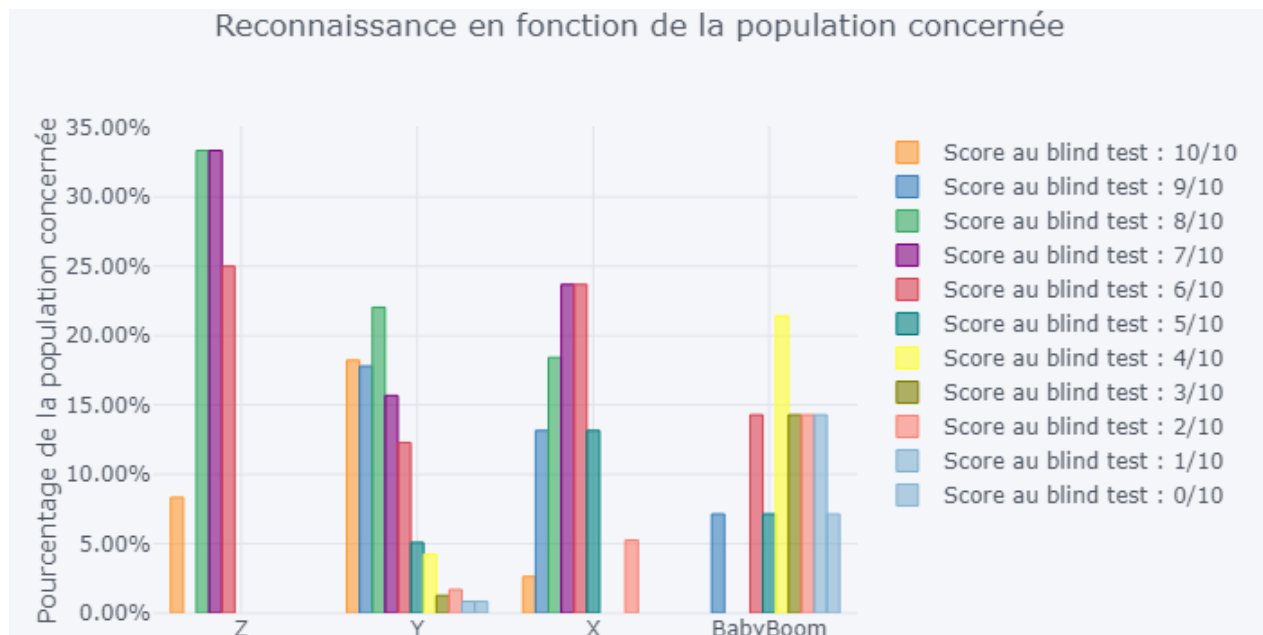


Figure 7: Score au blind test en fonction de la génération

Il y a 38 personnes dans la génération X, 236 dans la génération Y, 12 dans la génération Z et 14 dans la génération BabyBoomer.

On remarque que la génération Y est celle qui est la plus représentée. Cela introduit un fort biais pour le test suivant car la génération Z et la génération Babyboomer sont sous-représentées. Ainsi les tests de Normalité pour les distributions de la génération Z et Babyboom peuvent être remis en cause (annulant donc la possibilité d'ANOVA).



Nous continuons ici en supposant que l'ANOVA est possible.

2.6.2 Application des test

La moyenne de Z est de 0.73. La moyenne de Y est de 0.75. La moyenne de X est de 0.68. La moyenne de BabyBoomer est de 0.36.

Après application du test, nous obtenons les valeurs suivantes : $pvalue = 2.073e^{-10}$, $t = 17.36$. Donc $pvalue < 0.05$.

Bien que les échantillons soient peu représentatifs de toutes les générations, le test ANOVA infirme l'hypothèse de base. En effet la pvalue est bien inférieure à 0.05. Donc les moyennes sont différentes. On remarque que la moyenne des BabyBoomer est bien en deçà du reste.



Pour rappel l'échantillon des babyboomers est très peu représentatif.

On refait une ANOVA pour comparer les génération X, Y et Z. Après application du test, nous obtenons les valeurs suivantes : $pvalue = 0.0980$, $t = 2.341$. Donc $pvalue < 0.05$.



Ici la p-value est de 0.10, nous ne pouvons rejeter l'hypothèse nulle. Cependant il y a tout de même de fortes présomptions contre le modèle. Il est intéressant de noter que, les trois générations ont été marquées par ces films. Ils sont donc inter-générationnels, ce nous voulions puisque c'est un critère des films cultes.

Pour rappel, les conclusions tirées de ce test sont à relativiser car l'ANOVA ici est hautement faussé par la taille très faible de l'échantillon.

2.7 Hypothèse 7

Les personnes n'ayant pas vu le film reconnaissent aussi bien que celles qui l'ont vu.

► $H_0 : \mu_{\text{film_pas_vu_mais_reconnu}} = \mu_{\text{vu_film_et_reconnu}}$

► $H_a : \mu_{\text{film_pas_vu_mais_reconnu}} \neq \mu_{\text{vu_film_et_reconnu}}$

L'objectif est de comparer, film par film, si les gens n'ayant pas vu ce film peuvent en connaître les répliques et musiques aussi bien que ceux qui l'ont déjà regardé.

2.7.1 Visualisation

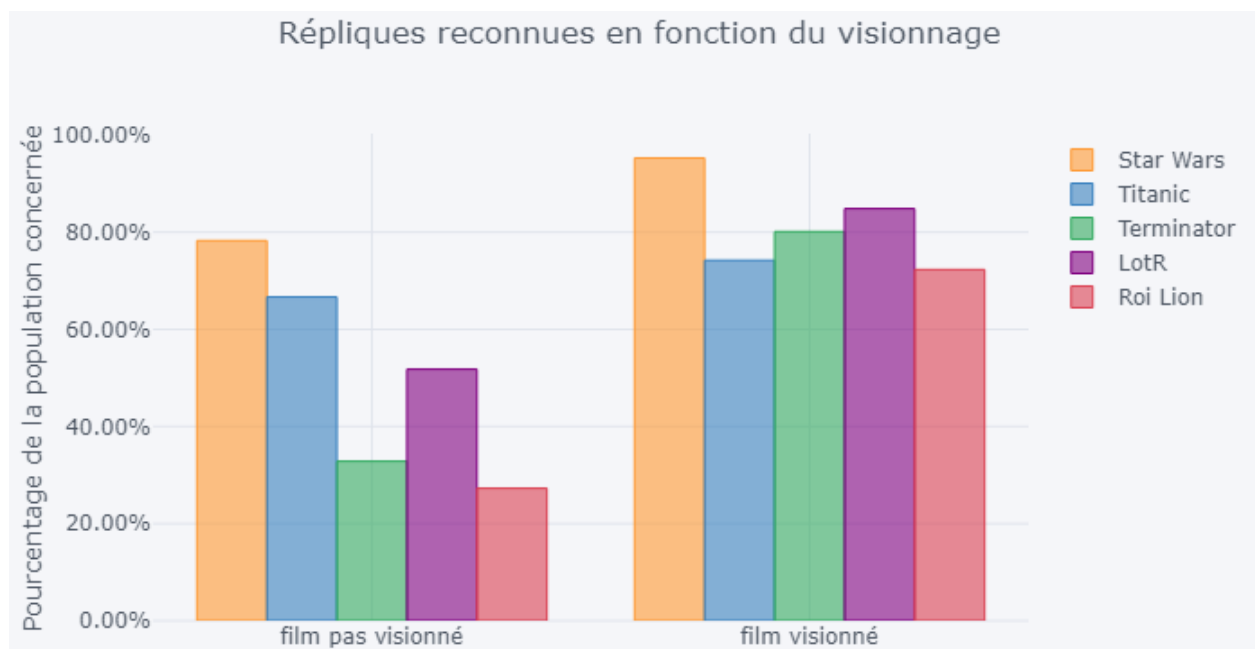


Figure 8: Reconnaissance de la musique sachant que le film a été visionné ou non

Comme prévu, tous les films semblent être mieux reconnus en répliques et musiques lorsque le film a été visionné. Cependant, cela semble varier d'un film à l'autre, nous allons le vérifier.

On peut remarquer avec ces deux graphiques, qu'en moyenne, ceux qui ont vu Terminator ont deux fois plus de chance de trouver la réponse au Blind Test que ceux qui ne l'ont pas vu. Ce film étant le moins populaire, cela nous conforte dans nos prédictions.

De plus, Titanic est le film le plus vu de tous les temps au box-office français et Star Wars est extrêmement populaire. Il est normal que les personnes ne les ayant pas vu en aient été fortement imprégnées par leur entourage et donc que les meilleures scores des films non visionnés soient pour ces deux films.

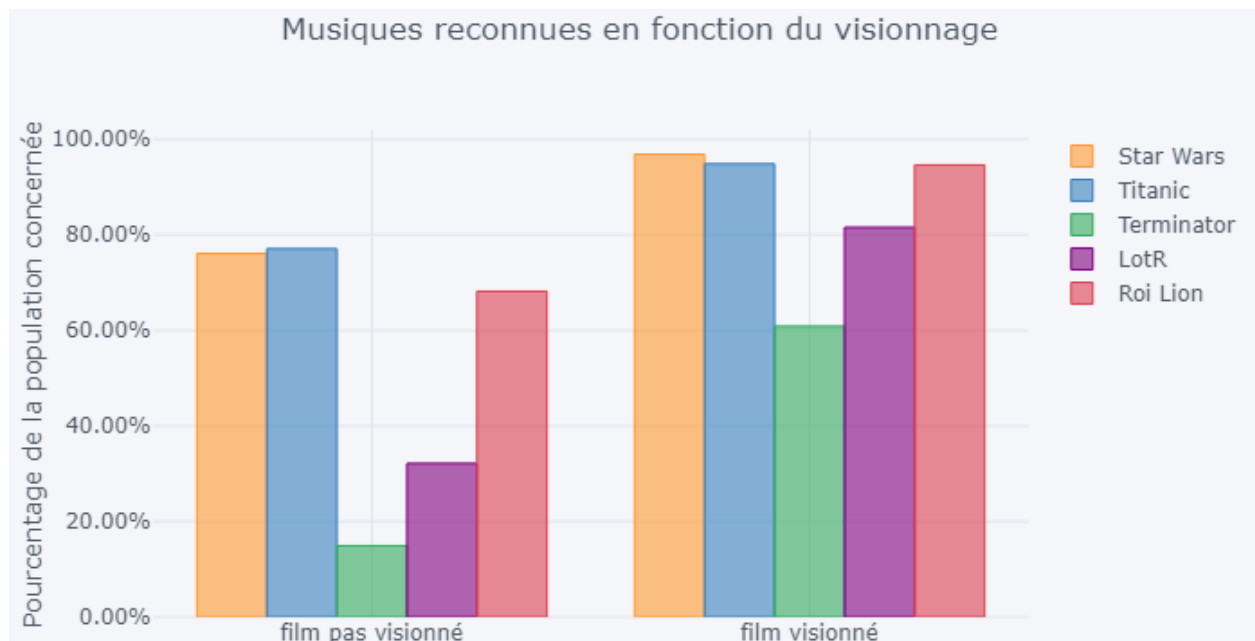


Figure 9: Reconnaissance de la réplique sachant que le film a été visionné ou non



Seulement 22 personnes sur les 300 interrogées n'avaient jamais vu le Roi Lion, ce qui perturbe fortement les mesures et leur donne une grande volatilité. Nous pourrions remettre en cause la viabilité d'un test de Student. Cependant, vu le nombre de personnes n'ayant pas vu les 4 autres films, cette valeur est acceptable.

2.7.2 Application des test

Nous appliquons un test t de Student apparié pour chaque film.

2.7.2.1 H7.1: Les personnes n'ayant pas vu Star Wars reconnaissent aussi bien que celles qui l'ont vu.

La moyenne de pourcentage de reconnaissance du film par ceux qui ne l'ont pas vu est de 0.75 et la moyenne de pourcentage de reconnaissance du film par ceux qui l'ont vu est de 0.95

Après application du test, nous obtenons les valeurs suivantes : $pvalue = 1.143e^{-08}$, $t = -5.872$. L'intervalle de confiance des différence de moyenne est de : [0.10,0.31].



$pvalue \ll 5\%$ donc on peut rejeter l'hypothèse nulle. Les personnes n'ayant pas vu Star Wars reconnaissent moins bien que celles qui l'ont vu.

2.7.2.2 H7.2: Les personnes n'ayant pas vu le Roi Lion reconnaissent aussi bien que celles qui l'ont vu.

La moyenne de pourcentage de reconnaissance du film par ceux qui ne l'ont pas vu est de 0.39 et la moyenne de pourcentage de reconnaissance du film par ceux qui l'ont vu est de 0.79

Après application du test, nous obtenons les valeurs suivantes : $pvalue = 1.805e^{-08}$, $t = -5.7877$. L'intervalle de confiance des différence de moyenne est de : [0.23,0.59].



Les personnes n'ayant pas vu le Roi Lion reconnaissent moins bien que celles qui l'ont vu.

2.7.2.3 H7.3: Les personnes n'ayant pas vu Terminator reconnaissent aussi bien que celles qui l'ont vu. La moyenne de pourcentage de reconnaissance du film par ceux qui ne l'ont pas vu est de 0.18 et la moyenne de pourcentage de reconnaissance du film par ceux qui l'ont vu est de 0.64

Après application du test, nous obtenons les valeurs suivantes : $pvalue = 9.7334e^{-08}$, $t = 12.395$. L'intervalle de confiance des différence de moyenne est de : [0.40,0.54].



Les personnes n'ayant pas vu Terminator reconnaissent moins bien que celles qui l'ont vu.

2.7.2.4 H7.4: Les personnes n'ayant pas vu Titanic reconnaissent aussi bien que celles qui l'ont vu.

La moyenne de pourcentage de reconnaissance du film par ceux qui ne l'ont pas vu est de 0.68 et la moyenne de pourcentage de reconnaissance du film par ceux qui l'ont vu est de 0.81

Après application du test, nous obtenons les valeurs suivantes : $pvalue = 0.0036$, $t = -2.9335$. L'intervalle de confiance des différence de moyenne est de : [0.03,0.25].



Les personnes n'ayant pas vu Titanic reconnaissent moins bien que celles qui l'ont vu.

2.7.2.5 H7.5: Les personnes n'ayant pas vu le Seigneur des anneaux reconnaissent aussi bien que celles qui l'ont vu.

La moyenne de pourcentage de reconnaissance du film par ceux qui ne l'ont pas vu est de 0.34 et la moyenne de pourcentage de reconnaissance du film par ceux qui l'ont vu est de 0.79

Après application du test, nous obtenons les valeurs suivantes : $pvalue = 3.6178e^{-16}$, $t = -8.6347$. L'intervalle de confiance des différence de moyenne est de : [0.34,0.55].



Les personnes n'ayant pas vu le Seigneur des Anneaux reconnaissent moins bien que celles qui l'ont vu.

2.7.3 Conclusion

On remarque que toutes les sous-hypothèses de l'hypothèse 6 ont une p-value très faible, et bien inférieure à 0.05. Nous pouvons donc rejeter l'hypothèse nulle de toutes les hypothèses. Ainsi, comme nous pouvions nous en douter, les personnes ayant vu le film reconnaissent mieux le film lors du blind test comparé à ceux qui ne l'ont jamais visionné. En cohérence avec ce qui a été dit plus tôt, Terminator a bien la p-value la plus faible (E-29) et Titanic la plus élevée (0.3%).

2.8 Hypothèse 8

H7: Un film vu jeune a plus d'impact sur la mémoire (reconnaissance de réplique OU musique du film au blind test) qu'un autre.

➤ $H_0 : \mu_{0-10} = \mu_{11-20} = \mu_{>20}$

➤ $H_a : \text{Il y a une différence de moyenne}$

Nous essayons de voir ici si l'âge de premier visionnage a un impact sur la mémorisation d'un film.

2.8.1 Visualisation

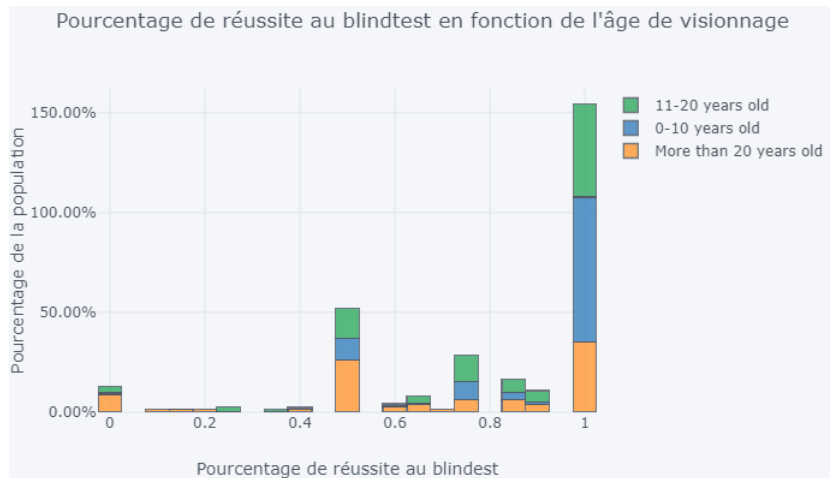


Figure 10: Pourcentage de réussite au blindtest en fonction de l'âge de visionnage

On remarque qu'une tendance se dégage: pour ceux qui ont vu le film avant leur 10 ans (rectangles bleus sur le graphe), ils ont tous plus de 50% de réussite au blind test.

2.8.2 Application des test

Il y a 77 échantillon dans la catégorie More than 20 years old

Il y a 212 échantillon dans la catégorie 0-10 years old

Il y a 268 échantillon dans la catégorie 11-20 years old



On peut remarquer que le nombre d'échantillon est cette fois-ci suffisant pour appliquer un test ANOVA.



IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

IMT Atlantique Bretagne - Pays de la Loire - www.imt-atlantique.fr

Campus de Brest
Technopôle Brest-Iroise
CS 83818
29238 Brest Cedex 03
T +33 (0)2 29 00 11 11
F +33 (0)2 29 00 10 00

Campus de Nantes
4, rue Alfred Kastler - La Chantrerie
CS 20722
44307 Nantes Cedex 03
T +33 (0)2 51 85 81 00
F +33 (0)2 51 85 81 99

Campus de Rennes
2, rue de la Châtaigneraie
CS 17607
35576 Cesson Sévigné Cedex
T +33 (0)2 99 12 70 00
F +33 (0)2 99 12 70 08