INSA
ROUEN NORMANDIE

Étude statistiques des données d'un speed dating

STPI 2, M8

Auteurs :
Mathilde GALLAND

Maxime PRIEUR

Enseignant :
Stéphane CANU

Introduction

Dans le cadre de notre EC de M8, Introduction à la science des données, nous avons pu travailler sur des données concernant le speed dating. Nous avons choisis ce sujet après plusieurs recherches pour le nombre de données qui étaient à disposition sur ce sujet ainsi que l'intérêt au quotidien que pourrait contenir son analyse . La communcation déterminent les différentes affinités entre les êtres humains.

En effet, une grande majorité d'entre nous rêve un jour de trouver l'amour afin de fonder une famille et passer le reste de sa vie aux côtés de quelqu'un. Mais l'amour est-il vraiment aveugle ? N'y a-t-il pas des facteurs d'ordres professionnels, culturels ou encore éthiques qui font que les gens tombent amoureux et s'attirent. L'objectif principal de ce projet est de vérifier si les critères pris en compte et donnés au départ par les hommes et les femmes sont bien en accord avec leurs choix.

Grâce à la multitude de données fournies et après votre lecture de notre dossier sur l'analyse que nous en avons faites, il est plus que probable que nous ne repartirez dorénavant plus seul de vos prochains speed datings.

Table des matières

1 Description des variables

Les données fournies représentent 8379 observations d'individus résidant à Columbia (ville située en Caroline du sud aux Etats-unis), associées à différentes variables.

Nous n'avons sélectionné qu'une seule partie des variables qui ont un intérêt pour notre étude.

1.1 Un peu de vocabulaire sur le speed-dating

Le speed dating est une méthode de recherche sentimentale qui consiste en une série d'entretiens courts avec différents partenaires potentiels. A l'issu de chaque rendez-vous appelez « dates », les personnes célibataires sont invitées à mettre une appréciation en fonction de plusieurs critères précis. Si les personnes se plaisent, alors on dit qu'il y a match.

Maintenant que vous connaissez tout le vocabulaire technique, nous allons vous expliquez quels critères il faudra mettre en avant pour plaire à coup sûr à votre future dulcinée.

1.2 Les matchs et ratio de match

La variable match est la plus importante de nos données, celle-ci permettant de déterminer si les personnes se plaisent de façon réciproque ou non, en d'autre termes, cette variable sera la variable la plus importante à expliquer.

Match est une variable qualitative. En effet, la valeur est de 1 quand il y a un match entre les deux personnes et 0 quand ce n'est pas le cas.

Match	Non Match
1070	5557

Puisque cette variable est qualitative (binaire) et que les individus participent à plusieurs dates on crée la variable ratio de match en sommant le nombre de matchs par personnes et divisant par le nombre de dates. Ce programme nous a aidé à faire ce ratio :

```
%% Faire les ratios de matchs par personnes
%%

j=1;
V(:,1)= M(1,:)
compteur=1;
for i=2:759 % Taille de M
    if M(i,1)==j
       V(j,[2 3 4 5 6 7 8])=M(i,[2 3 4 5 6 7 8])+V(j,[2 3 4 5 6 7 8]);
       compteur=compteur+1;
    else
        V(j,[2 3 4 5 6 7 8])=V(j,[2 3 4 5 6 7 8])/compteur;
        j=j+1;
        V(j,:)=M(i,:);
        compteur=1;
    end

end
```

FIGURE 1 – Calcul des différents ratios

On trouve rapidement sur Matlab un ratio de match moyen de 0.1682 et une médiane de 0.1364.

Afin d'y voir plus claire, rien de mieux qu'une boite à moustache et le graphique de la fonction de répartition empirique.

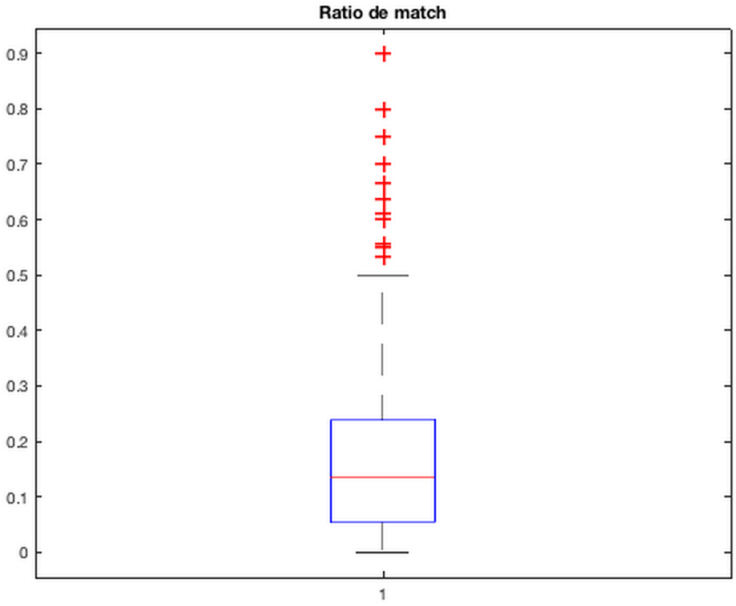


FIGURE 2 – Boîte à Moustache du ratio de Match

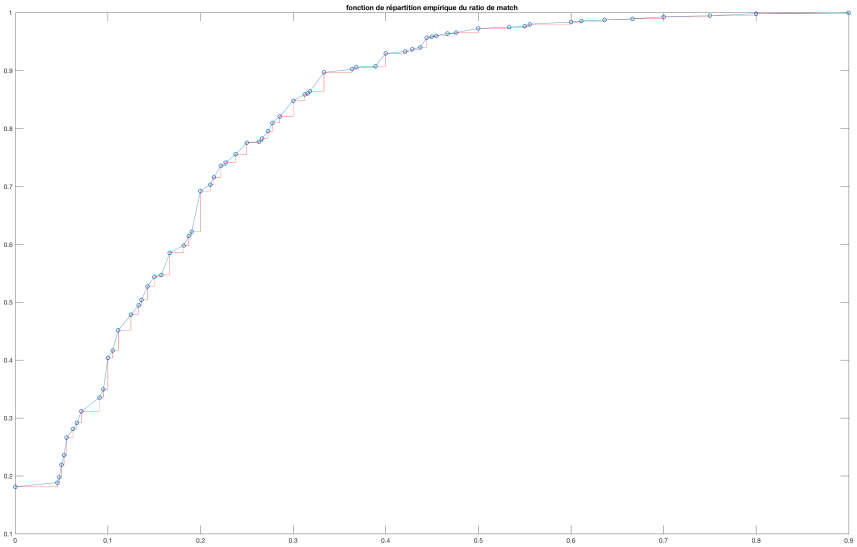


FIGURE 3 – Fonction de répartition empirique du Ratio de Match

Comme nous le montre la boîte à moustache, la majorité des individus peuvent estimer matcher avec une personne rencontrée avec une chance (chance vraiment ? nous le saurons plus tard) entre 6 et 24%. Toutefois on peut remarquer que certaines personnes sont de vrais séducteurs et arrivent quand à eux à matcher avec 90% des individus rencontrés.

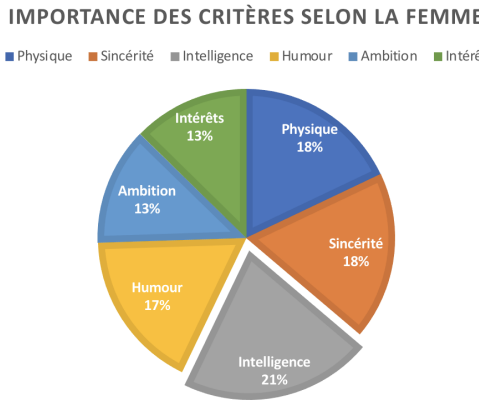
1.3 Les critères de choix du partenaire

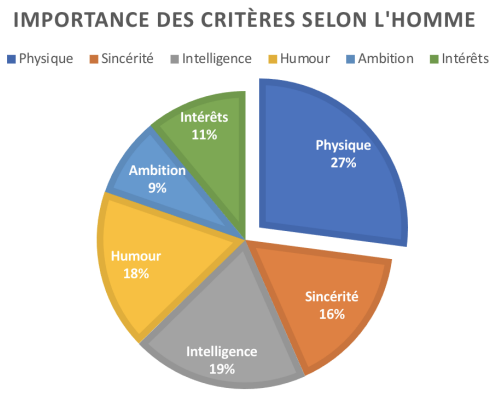
Les personnes du speed dating sont jugées en fonction de différents critères (notes attribuées entre 0 et 10) :

- l'attirance physique (variable attr)
- la sincérité (variable sinc)
- l'intelligence (variable intel)
- l'humour (variable fun)
- l'ambition (variable amb)
- Les intérêts communs (variable shar)

L'objectif est de comparer ces différents critères et de déterminer l'importance de ces derniers dans le choix des partenaires.

Nous pousserons l'étude si possible en nous demandant quels sont les critères les plus importants pour les hommes et pour les femmes.





Ces deux diagrammes nous montrent que les femmes et les hommes ne recherchent pas la même chose. Alors que les hommes restent focalisés sur l'apparence, les femmes portent plus d'intérêts à l'intelligence. Dans la suite de notre projet, nous avons cherché à vérifier si ces affirmations sont réellement appliquées ou non.

2 Analyse en composantes principales

Nous réalisons notre ACP sur le tableau de l'ensemble des individus (552) en prenant en compte les 6 critères de choix de partenaire (variables quantitatives).

Afin de comparer les données entre elles, bien que l'effet d'échelle soit limité ici (les variables étant toutes des notes sur 10), il est nécessaire de centrer réduire nos données. On prend donc notre matrice X avec n=552 le nombre d'individus/lignes et p=6 le nombre de colonnes/variables. On calcule Xnnotre matrice centrée réduite à l'aide de $\overline{X} = ones(n,1) * mean(X)$,la moyenne, et$S = ones(n,1) * std(X)$, l'écart type.

$$Xn = \frac{X - \overline{X}}{S}$$

On calcule ensuite la matrice de corrélation $C = \frac{X_T^n * X_n}{n}$ où chaque éléments $\alpha_{i,j}$ représente la corrélation comprise entre 0 et 1 de la variable i avec la variable j.

Si on s'intéresse à cette matrice C, on remarque une forte corrélation entre l'intelligence et la sincérité (0.6522), mais aussi l'intelligence et l'ambition et (0.6586). La plus grande corrélation reste entre l'humour et les intérêts communs (0.7602) bien que l'humour soit aussi relativement proche de l'attirance physique (0.6762).

Après quelques calculs sur matlab, nous obtenons les valeurs propres et vecteurs propres suivants :

```
Vp =

    0.1906
    0.2781
    0.3363
    0.6265
    1.1043
    3.4642
```

FIGURE 4 – Valeurs propres

```
V =

   -0.0097   -0.0669    0.7652   -0.1312    0.4912    0.3892
    0.3269    0.3273   -0.0285   -0.7071   -0.3859    0.3691
   -0.5555   -0.4804    0.0979    0.0198   -0.5399    0.3991
   -0.5186    0.4788   -0.4165    0.0332    0.3611    0.4437
    0.3426    0.3502    0.1639    0.6922   -0.3227    0.3872
    0.4452   -0.5529   -0.4514    0.0472    0.2895    0.4542
```

FIGURE 5 – Vecteurs propres

En ce qui concerne l'importance des valeurs propres attachées à chaque critère, on observe que 58% de l'information peut être résumée à partir d'une seule variable et 76% en deux variables.

Valeurs propres	Indice de qualité
3.4642	0.577366
1.1043	0.184054
0.6265	0.104415
0.3363	0.056044
0.2781	0.046354
0.1906	0.031766

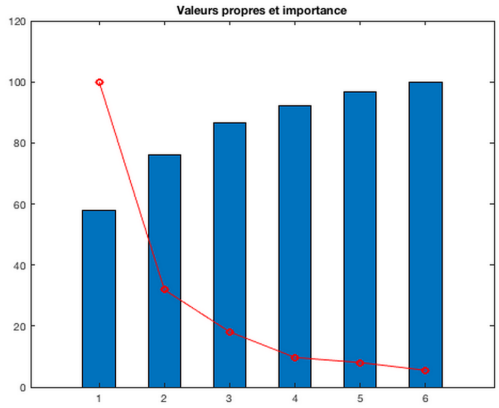
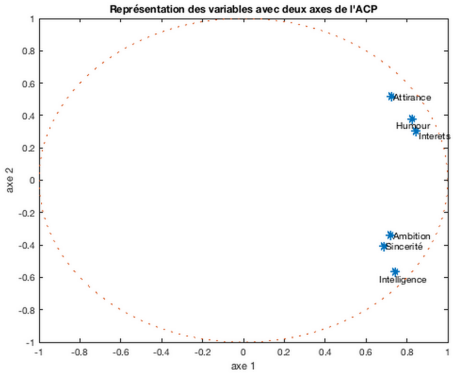


FIGURE 6 – Valeurs et importance

Le graphique représentant le cercle de corrélation et les variables sur le plan des 2 facteurs principaux nous confirme ce que nous a appris la matrice de corrélation. On discerne ici deux groupes, l'humour, l'attirance physique et les intérêts communs intimement liés et un groupe comprenant la sincérité l'ambition et l'intelligence. Il est intéressant de remarquer que l'attirance et l'intelligence sont les deux critère les moins corrélés selon les deux axes les plus importants.



3 Régression linéaire

3.1 Régression multiple (ensemble des variables)

Le but de cette partie est de vérifier si le ratio de match d'une personne peut être déterminé en fonction des moyennes des notes qui lui sont attribuées pour chacun de ses critères et quel critère influe le plus sur le ratio de match.

On pose tout d'abord le modèle suivant :

$$RatioMatch = a_0 + a_1 \cdot Attirance + a_2 \cdot Sincerite + a_3 \cdot Intelligence + a_4 \cdot Humour + a_5 \cdot Ambition + a6 \cdot Intérêt$$

en version matricielle : $y = X \cdot a$

Avec : $RatioMatch' = (RatioMatch_1 ... RatioMatch_n) = y' = (y_1 ... y_n)$; $a' = (a_1 ... a_n)$

$$\begin{pmatrix} 1 & Attirance_1 & Sincérité_1 & Intelligence_1 & Humour_1 & Ambition_1 & Interets_1 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & Attirance_n & Sincérité_n & Intelligence_n & Humour_n & Ambition_n & Interets_n \end{pmatrix}$$

FIGURE 7 –

Pour mener a bien notre régression sur matlab, nous utilisons le code suivant :

```
%% Regression linéaire

%%  Regression de la variable y en fonction des autres
y=Matrice(:,3)
X=[ones(size(y)) Matrice(:,[5 6 7 8 9])]
a = (X'*X)\(X'*y)
%% Résidus de la regression
e = y-X*a;

%%
SCT = sum((y-mean(y)).^2);
SCM = sum((X*a-mean(y)).^2);
SCR = e'*e;

%% Qualité de la régression

R2 = 1 - SCR/SCT
```

FIGURE 8 – Code de la régression linéaire

On obtient d'après notre script :

$a' = (-0.31860.0274 - 0.00690.01310.0227 - 0.00350.0281)$et $R^2 = 0.2312$

Le coefficient de régression trouvé précédemment est très inférieur à 1 ce qui n'est pas suffisant pour la qualité de notre régression. Nous allons donc tester chaques variables séparément.

3.2 Régressions multiples spécifiques

3.2.1 R^2 de chaque variable

Variable	Modèle	R^2
Attirance	$RationMatch = a \cdot Attirance + b$	0.1823
Sincérité	$RationMatch = a \cdot Sincerite + b$	0.0411
Intelligent	$RationMatch = a \cdot Intelligence + b$	0.0481
Humour	$RationMatch = a \cdot Humour + b$	0.1808
Ambition	$RationMatch = a \cdot Ambition + b$	0.0524
Intérêts	$RationMatch = a \cdot Interets + b$	0.1868

Les intérêts communs suivis de près par l'attirance physique ont tous deux les plus gros R^2.

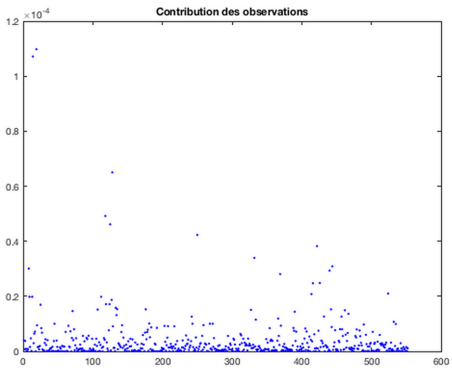
3.2.2 Régression

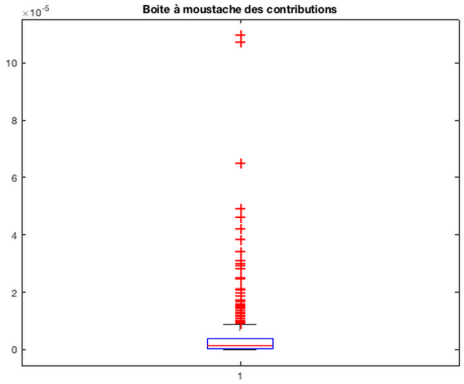
Variable	Modèle	R^2
Attirance	$RationMatch = a0 + a1 \cdot Attirance$	0.1823
Attirance+Intérêts	$RM = a0 + a1 \cdot Attirance + a2 \cdot Intérêts$	0.2226
Attirance+Intérêts+Hum	$RM = a0 + a1 \cdot Att + a2 \cdot Intérêts + a3 \cdot Hum$	0.2301
Att+Int+Hum+Amb	$RM = a0 + a1 \cdot Att + a2 \cdot Int + a3 \cdot Hum + a4 \cdot Amb$	0.2301
Att+Int+Hum+Amb+Intel	$RM = a0 + a1 \cdot Att + a2 \cdot Int + a3 \cdot Hum + a4 \cdot Amb + a5 \cdot Intel$	0.2307
Att+Int+Hum+Amb+Intel+Sin	$RM = a0 + a1 \cdot Att + a2 \cdot Int + a3 \cdot Hum + a4 \cdot Amb + a5 \cdot Intel + a6 \cdot Sinc$	0.2312

Puisque notre R^2 est faible, il convient de vérifier l'existence de points aberrants et penser à les retirer. Le R^2 n'augmente pas à l'ajout de la variable ambition, et augmentent peu pour l'ajout des critères suivant.

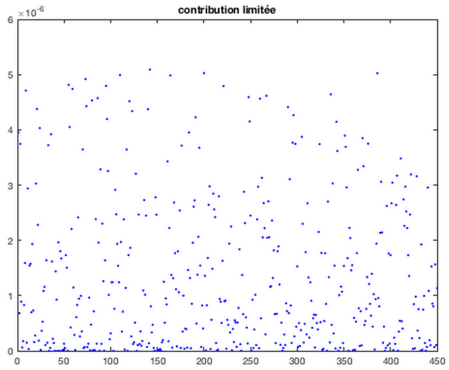
3.3 Régression sans points aberrants

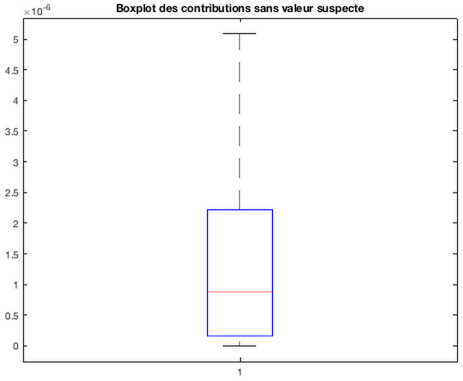
On trace tout d'abord la contribution de chaques observations. En traçant une boite à moustache des contributions, on note que les contribution sortent de l'épure pour une valeur au dessus de $8.94 \cdot 10^{-6}$.





On supprime par la suite les observations ayant une contribution supérieure à l'épure pour n'avoir plus aucune valeur suspecte.





Cette suppression de valeur contribuant de façon trop importante ne permet d'élever notre R^2 qu'à une valeur de 0.3656 ce qui reste très en dessous d'une régression linéaire de bonne qualitée.

3.4 Aparté , variable la plus explicative selon le genre

Pour regarder quelle est la variable la plus explicative du ratio de match , il est possible de s'intéresser aux coefficients de régression standardisés.

```
%% Variable la plus explicative
Coeff=a(2:6);
Predicteurs=Matrice(:,[5 6 7 8 9]);
Coefstd=(std(Predicteurs)/std(y)).*Coeff';
[maxvalue index]=max(abs(Coefstd));% Index contient le numéro de la variable recherchée
disp(index)
```

D'après le programme Matlab précédent, l'attirance physique serait la variable la plus explicative quant au résultat du match tous genres confondus.

3.5 Le plus important pour l'homme

De la même façon on trouve un $R^2 = 0.1797$, un $R^2 = 0.2430$et des critères classés de la façon suivante :

Attirance physique	Intérêts communs	Humour	Intelligence	Sincérité	Ambition

En séparant les deux genres, on s'aperçoit que la qualité de la régression se trouve améliorée. Et comme le montre le critère qui diffère en deuxième position dans le choix du match, les deux genres n'ont pas la même façon de sélectionner leur partenaire. On remarque que le classement des critères selon la femme ne sont pas les mêmes entre ce qui est dit avant les matchs et selon comment sont sélectionnés les matchs.

3.6 Le plus important pour la femme

En reprenant les étapes faites précédemment, nous trouvons un $R^2 = 0.2937$et $R^2 = 0.5352$ en limitant les contributions trop importante. En standardisant les coefficients de régression, on classe l'importance des critères pour le sexe féminin de la façon suivante :

Attirance physique	Humour	Intérêts communs	Intelligence	Sincérité	Ambition

4 Test de Student

Nous alons maintenant réalisé des test de Student afin de savoir si le match est indépendant de chacun de critères. Pour cela nous avons encore une fois transformé notre variable match de base qualitative en variable quantitative à l'aide du ratio.

L'objectif d'avoir des ratio est de faire une regression linéaire de chaque critère en fonction du match et de regarder, à l'aide de la p-valeur, si les critères sont ou non indépendants du Match.

4.1 Hypothèses

Nous avons défini les Hypothèses:

H0: Le match et le critère sont indépendants

H1: Le match et le critères sont dépendants

4.2 Modèle

Nous avons donc réalisé la regression linéaire avant de faire notre test de Student sur a.

La variance pour le test de Student est inconnue, d'où la formule suivante :

$$\hat{\sigma}^2 = \frac{1}{nt + np - 2}\left(\sum (x_{ti} - x_t)^2 + np \sum (x_{pi} - x_p)^2\right)$$

$$t = \frac{x_t - x_p}{\sqrt{\hat{\sigma}^2\left(\frac{1}{nt} + \frac{1}{np}\right)}}$$

On calculera ensuite la p-valeur telle que: $pval = P(T \leq t)$

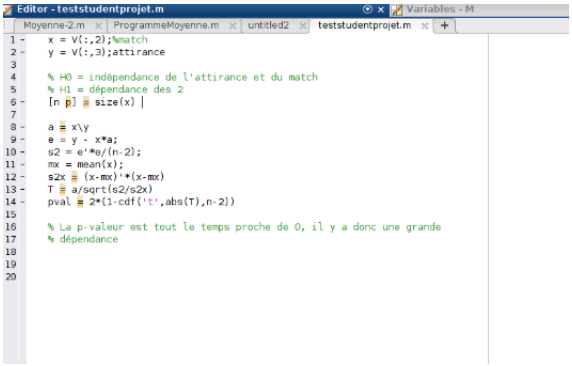


FIGURE 9 – Regression linéaire et p-valeur

4.3 Résultats

Critères	P-valeur
Attirance	0.00253
Sincérité	0.00645
Intelligence	0.00578
Humour	0.00337
Ambition	0.00978

La p-valeur est toujous inférieure à 0.05, donc on accepte H1. Notre résultat est logique car les critères sont forcément liés à la décision du match. Nous remarquons en plus que la p-valeur de l'attirance est plus faible que les autres, c'est donc le critère qui a le plus de dépendance avec le match.

5 Test du Chi 2

Ce test va nous permettre au travers d'un test d'indépendance, de vérifier la relation entre le ratio de match et les différents critères. Le test utilisant des variables qualitatives, il est nécessaire de transformer ces notes sur 10. Pour les cinq critères, on établit plusieurs niveaux, très faible (0-2.5), faible (2.5-5), moyen (5-7.5) et élevé (7.5-10).

5.1 Hypothèses

On pose tout d'abord les deux hypothèses :

H_0 : Le critère et le résultat du match n'ont pas d'influence l'un sur l'autre.

H_1 : Le critère et le résultat du match on un effet l'un sur l'autre.

5.2 Tableaux

Attirance	Très faible	Faible	Moyen	Elevé	Total
Match	9	42	571	591	1213
Non Match	299	937	3349	1233	5818
Total	308	979	3920	1824	7013

TABLE 1 – Tableau des observations

Attirance	Très faible	Faible	Moyen	Elevé	Total
Match	52.59	166.14	666.96	309.57	17%
Non Match	256.77	811.17	3256	1511.45	83%
Total	4.4%	13.9%	55.8%	25.9%	7.013

FIGURE 10 – Tableau Théorique

On calcule les marginales en bleu, $p_j = \frac{N_{\bullet j}}{n}$, ici n =7031 et les effectifs théoriques en rouge$p_j \cdot p_i \cdot n$.

5.3 Distance du χ^2

Il est ensuite nécessaire de calculer la distance du chi 2,

$$D(O,T) = \sum\sum(\frac{O_{i,j} - T_{i,j})^{\mathbf{2}}}{T_{i,j}} = 478.9412$$

Selon l'hypothèse H_0, D est distribuée d'après une loi du χ^2 à 3 degrés de liberté. En se référant à la table du χ^2 , la distance critique pour un risque admis de 0.001 est de 16.266. La valeure obtenue est bien supérieure à la valeur critique. On peut donc rejeter de façon nette et précise, l'hypothèse d'indépendance.

5.4 Résumé Test du χ^2 pour tous les critères

Critère	Distance du χ^2	Conclusion
Attirance	478.9412	Indépendance rejetée
Sincérité	883.2232	Indépendance rejetée
Intelligence	1050	Indépendance rejetée
Humour	697.8126	Indépendance rejetée
Ambition	459.6726	Indépendance rejetée

Ainsi les test du χ^2 effectués montreent bien que chaque critère a une influence sur le résultat du match.

6 Anova à un facteur

L'ANOVA ou analyse de la variance va nous permettre de vérifier que les intérêts communs influent sur le choix du partenaire. Pour ce faire, nous reprenons la variable match fournie originellement, c'est à dire la variable qualitative binaire, et nous l'opposons à la variable quantitative donnant la corrélation des intérêts entre les deux partenaires.

L'ANOVA se résume en un test d'égalité de la moyenne. Pour cela, on décompose la variance de Y en deux parties : - la Variances interclasses (différentes pour chaque groupes, ici ayant matché ou non). - la Variances intraclasses ou erreurs (attribuée aux variations aléatoires).

L'intensité de la liaison entre les deux variables est généralement mesurée par :

$$R^2 = \frac{\sum n_i(\mu_i - \mu)^2}{\sum (y_j - \mu)^2}$$

6.1 Poser les hypothèses

On pose d'abord deux hypothèses statistiques :

— la première H_0 : Les deux variables n'ont pas d'influence l'une sur l'autre.
— La seconde H_1 : les deux variables ont une influence l'une sur l'autre.

Afin d'effectuer l'anova sur Matlab, on trie nos données en séparant les individus ayant matché et ceux n'ayant pas réussis. On supprime un nombre d'observation non match pour en avoir autant que de match et avoir un plateau dit équilibré.

6.2 Décomposition de la variance

Considérons k groupes Y_k d'effectif n_k. Chaque individu s'écrit $y_{i,j}$ où ireprésente le groupe et j la place de l'individu dans son groupe.

on pose le modèle $y_{i,j} = a_i + \epsilon_{i,j}$

sois $SCE_{total} = SCE_{facteur} + SCE_{résidu}$ la somme des carrés des écart (SS en anglais), elle est exprimée par la variabilité inter-classe $SCE_{facteur} = \sum n_i(\overline{y_i} - \overline{y})^2$ et la variabilité intra-classe, $SCE_{résidu} = \sum\sum n_i(y_{ij} - \overline{y_i})^2$.

Dans notre cas : $y = 0.1842$ $y1 = 0.2173$ $y2 = 0.1511$

$SCE_{facteur} = 2.963$ $SCE_{résidu} = 260.4251$ $SCE = 263.3885$

6.3 Analyse des résidus

Pour pouvoir procéder à l'analyse de la variance entre deux variables, il est nécessaire de vérifier la condition que les résidus suivent une loi proche de la loi normale.

Par matlab, on calcule l'espérance des résidus, $\sum = -2.5 \cdot 10^{-16}$ et donc environ 0. Les résidus sont également répartis comme on peut le voir ci-dessous, en approximant une loi centrée réduite.

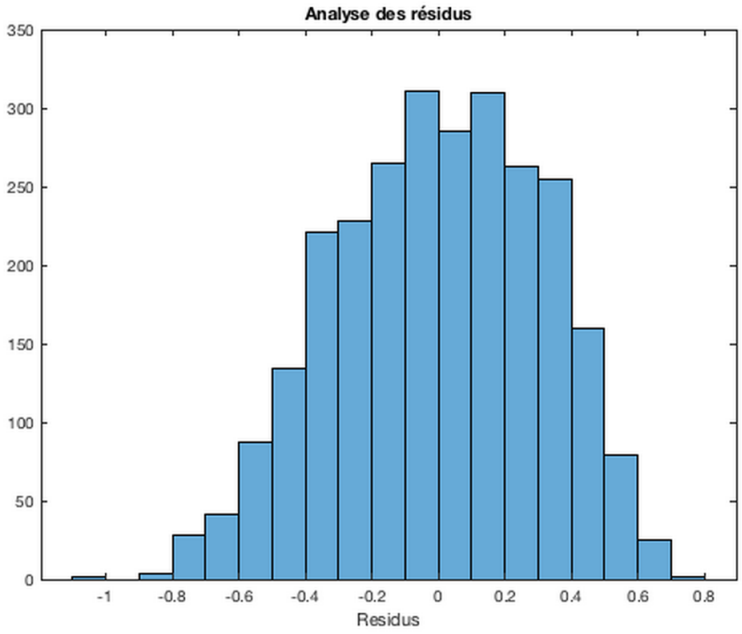


FIGURE 11 – Analyse des résidus

6.4 Test de Fisher

La loi de Fisher est définie comme le rapport de deux lois du 2. Or, la somme des carrés des écarts , SCE, suit une loi du 2. Dès lors, on pose les variances $S^2_{facteur} = \frac{SCE_{facteur}}{p-1}$ et $S^2_{résidu} = \frac{SCE_{résidu}}{n-p}$ (on divise les SCE par leur degré de liberté).

On peut ainsi poser $F = \frac{S^2_{facteur}}{S^2_{résidu}}$ variable suivant une loi de Fisher.

Dans notre cas : $S^2_{facteur} = 2.962$ $S^2_{résidu} = 0.0965$

$F = 30.7013$

On vérifie la justesse de nos calculs et les hypothèses par matlab avec la fonction anova1.

ANOVA Table

Source	SS	df	MS	F	Prob>F
Columns	2.963	1	2.96345	30.7	3.30062e-08
Error	260.425	2698	0.09653		
Total	263.389	2699			

Heureusement nos calculs sont correctes ! De plus puisque p-value est suffisamment faible, on peut rejeter l'hypothèse 0 , et dire que les intérêts communs entre deux partenaires ont bien une influence sur le résultat du match. Toutefois, $R^2 = 0.0113$ reste faible.

Conclusion

Ce projet nous aura permis d'apprendre à analyser des données pour répondre à une problématique précise. Dorénavant, nous pouvons dire, à travers notre analyse, que chaque critère que ce soit l'attirance physique, l'Intelligence, l'Humour, les Intérêts communs, l'Ambition ou encore la Sincérité a une importance pour plaire à une personne.

En comparant l'importance des critères des femmes et des hommes, nous avons remarqué que tout le monde a à peu près les mêmes critères de choix et que l'attirance physique prône devant toutes les autres.

Malgré tout, le speed dating n'est pas forcément représentatif de la réalité. Il est impossible d'apprendre à connaître une personne lors d'un entretien de 2 minutes, c'est pourquoi le physique est le premier critère pris en compte lors des matchs. En effet, en comparant l'importance des critères des hommes et des femmes, on remarque que l'intelligence est plus importante pour la femme alors que l'homme regardent plutôt l'apparence physique. Mais quand on compare avec la réalité des matchs. La plupart des match sout dus à l'Apparence Physique. Les femmes se voileraient-t-elle la face ?

Et n'oubliez quand même pas, les critères sont importants mais comme dirait Blaise Pascal : "Le coeur a ses raisons que la raison ignore".

Annexe

Provenance des données

Les données proviennent du site https ://data.world/annavmontoya/speed-dating-experiment. Un fichier contenant les données en format excel ainsi qu'un fichier word décrivant chaques variables sont disponible. L'ensemble de ces données ont été fournies par Ray Fisman et Sheena Iyengar deux professeurs de Columbia business school dans le cadre de leur travail intitulé "Gender Differences in Mate Selection : Evidence From a Speed Dating Experiment".

Ces données ont été collectées à l'occasion de Speed datings ayant eu lieux entre l'année 2002 et 2004. Dans un premiers temps se déroulait des rencontres entre chaques personnes d'une durée de 4 minutes. A la fin de chaque rendez vous les participants devaient noter leur partenaire selon les critères étudiés dans notre projet, et s'ils avaient été séduit par leur partenaire ou non. Ces données comportent également de nombreuses informations que nous n'avons pas étudiées dans notre projet tels que la richesse du quartier dans lequel le sujet a grandit, la manière dont il s'évalue par rapport au regard des autres, l'importance du choix de la religion chez la personne qu'il recherche...etc.

Lyengar et Fishman ont également réalisés une analyse intitulée "Racial Preferences in Dating". Cette analyse est disponible à l'adresse ci-dessous :
https ://faculty.chicagobooth.edu/emir.kamenica/documents/racialPreferences.pdf

Fiche sur les données

Les données récupérées contiennent 8379 observations pour 108 variables. voici les variables utilisées dans notre projet :

Variable	Description
iid	Numéro unique attribué au sujet
gender	binaire (1 pour un homme, 0 pour une femme)
match	binaire (1 pour un oui, 0 pour non)
wave	numéro de la session du speed dating
round	Nombre de partenaires rencontré au total dans la soirée
order	Moment de la soirée ou le partenaire a été rencontré
int_cor	Corrélation des intérêts avec le partenaire rencontré
pf_o : attr, sin, int, fun, amb, sha	Notes de l'individu sur ce qu'il recherche en face
attr, sinc, intel, fun, amb : _o	Notes que le partenaire a attribué au sujet