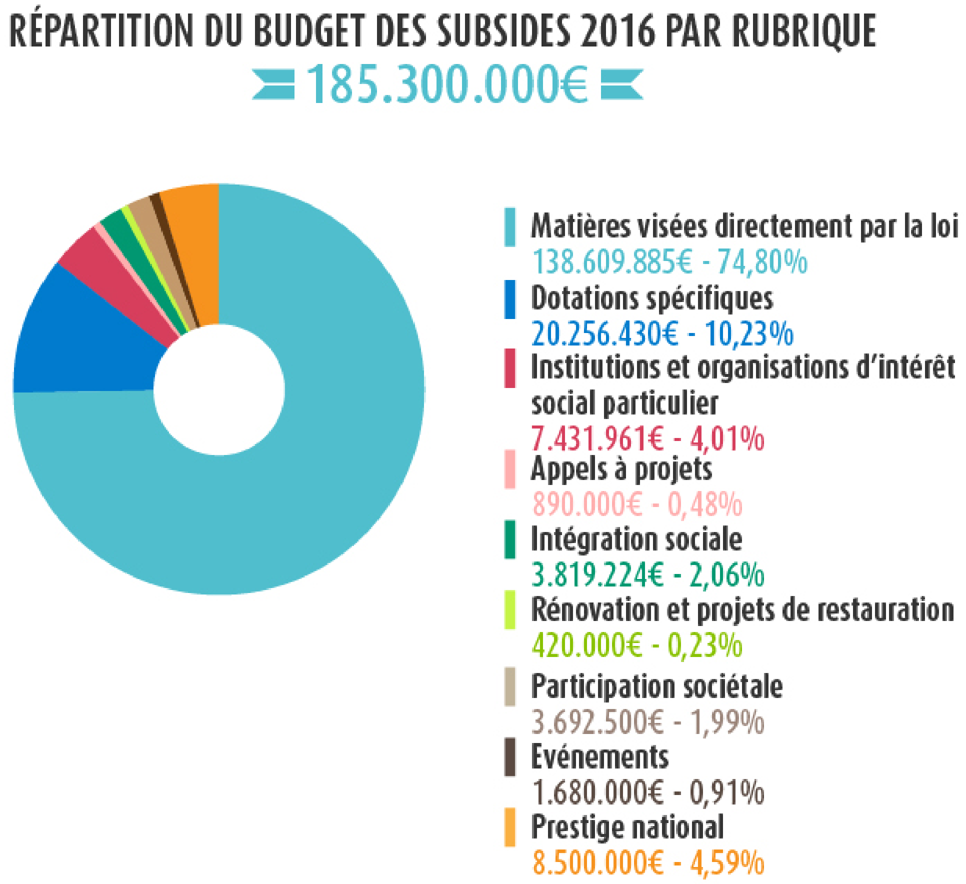
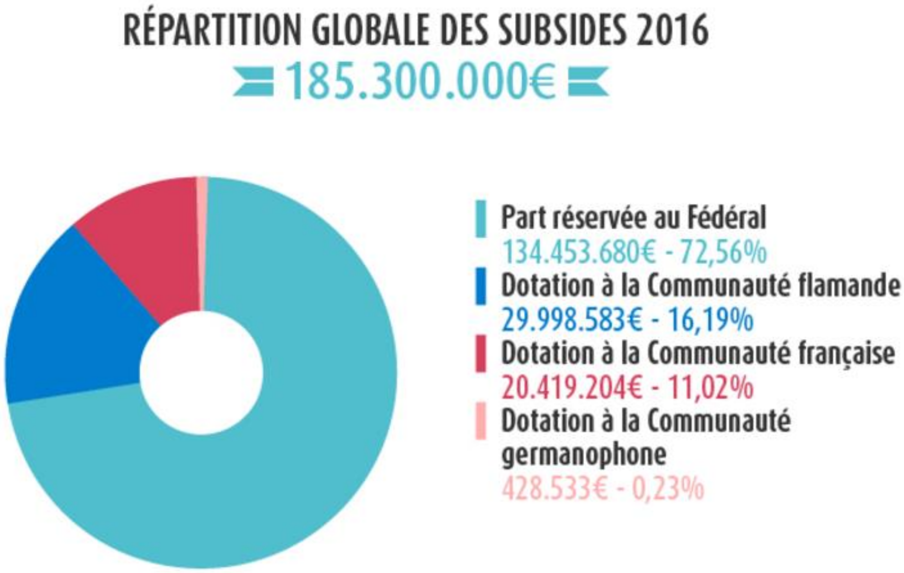
Annexes

Annexe 1 : Répartition des subsides et du sponsoring de la Loterie Nationale

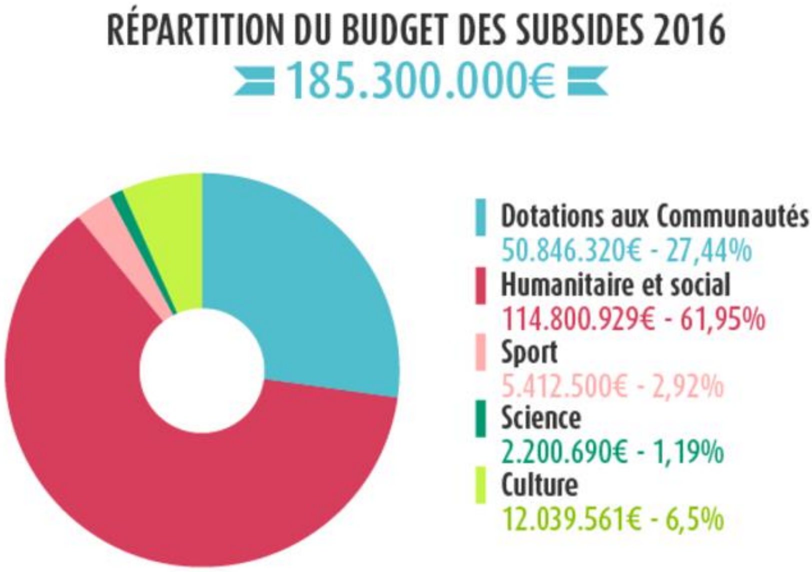
Voici quelques graphiques précisant la répartition des subsides et du sponsoring de la Loterie Nationale :



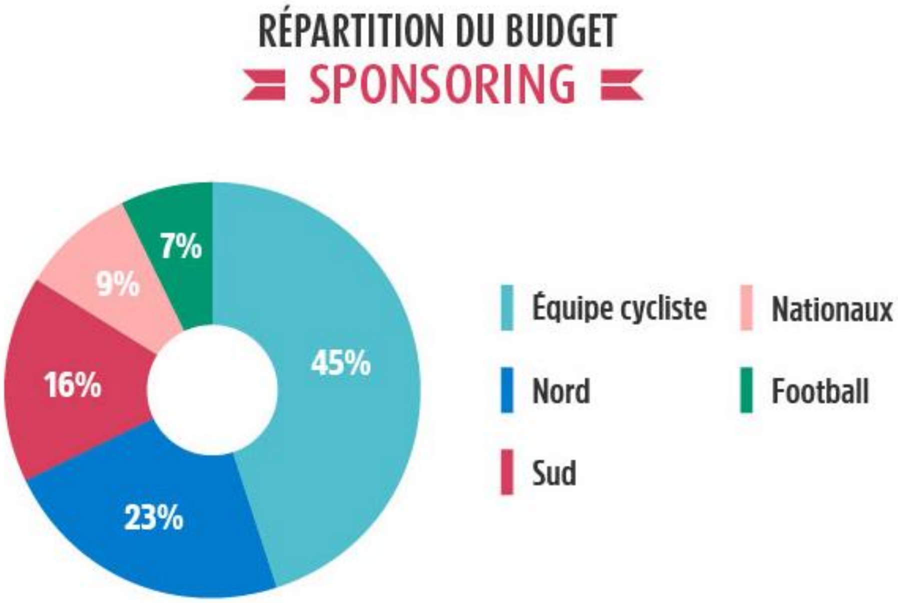
Source : Loterie Nationale



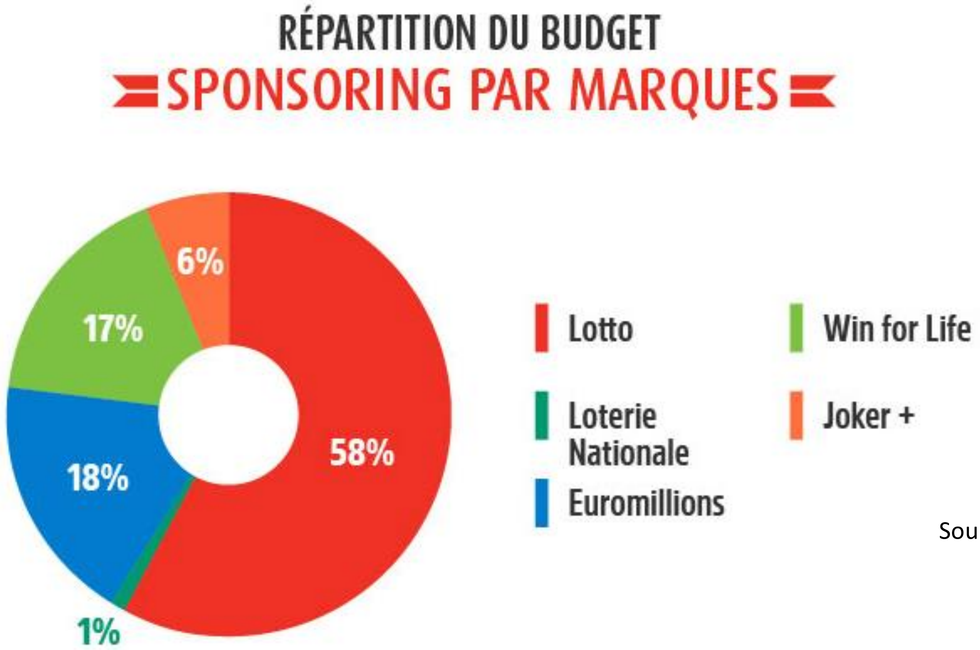
Source : Loterie Nationale



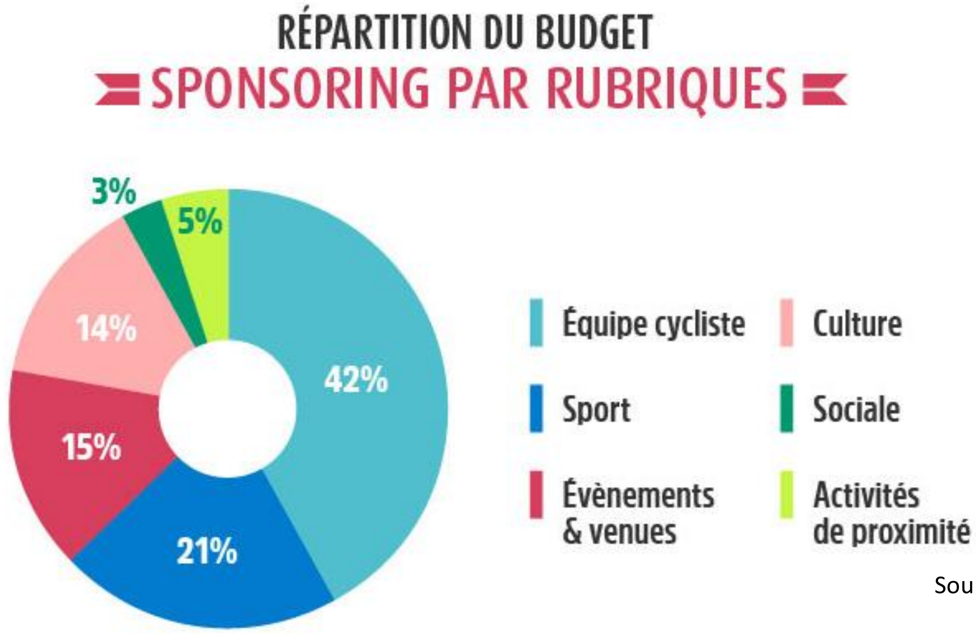
Source : Loterie Nationale



Source : Loterie Nationale



Source : Loterie Nationale



Source : Loterie Nationale

Annexe 2 : Le sondage

Loterie Nationale

Habitudes de jeu

1. Vous jouez généralement au(x) jeu(x) de tirage, pari(s) sportif(s) et e-game(s) suivant(s) (plusieurs réponses possibles)

- [ ] Je ne joue jamais aux jeux suivants

2. Vous jouez généralement au(x) jeu(x) de grattage suivant(s) (plusieurs réponses possibles)

- [ ] Je ne joue jamais aux jeux suivants

1

Loterie Nationale

Habitudes de jeu

3. A quelle fréquence jouez-vous aux jeux de grattage (WinForLife, Subito!, Presto, Cash, etc.)

- ◯ Tous les jours
- ◯ 2-3 fois par semaine
- ◯ 1 fois par semaine
- ◯ 1 fois tous les 15 jours
- ◯ 1 fois par mois
- ◯ Moins d'une fois par mois

4. Combien dépensez-vous en général lorsque vous achetez un ou des tickets à gratter (WinForLife, Subito!, Presto, Cash, etc.) ?

- ◯ 0 - 5 €
- ◯ 5 - 10 €
- ◯ 10-25 €
- ◯ 25 -50 €
- ◯ 50 - 100 €
- ◯ 100 € et plus

5. Combien dépensez-vous par mois en jeux de grattage ?

- ◯ 0€ - 5€
- ◯ 6€ - 10€
- ◯ 11€ - 15€
- ◯ 16€ - 20€
- ◯ 21€ - 25€
- ◯ 26€ - 35€
- ◯ 36€ - 50€
- ◯ 51€ - 75€
- ◯ 76€ - 100€
- ◯ 101€ - 150€
- ◯ 151€ - 200€
- ◯ 201€ et plus

6. Qu'est ce qui influence le plus votre choix lorsque vous achetez un ticket à gratter ? (Plusieurs réponses possibles)

- ☐ Le montant du gros lot
- ☐ La probabilité de gagner le gros lot
- ☐ Le nombre de lots intermédiaires
- ☐ La probabilité de gagner les lots intermédiaires
- ☐ Le prix du ticket
- ☐ La notoriété du ticket
- ☐ Le design du ticket
- ☐ Autre (veuillez préciser)

7. Pourquoi jouez-vous aux jeux de grattage ? (plusieurs réponses possibles)

- ☐ C'est plus par impulsion qu'autre chose
- ☐ Juste pour le plaisir
- ☐ Pour avoir une chance de gagner le gros lot
- ☐ C'est un passe-temps entre amis/collègues/famille
- ☐ Puisque la Loterie Nationale reverse une partie de ses bénéfices, c'est une bonne manière de lier l'utile à l'agréable
- ☐ Autre (veuillez préciser)

Loterie Nationale

Probabilités

8. De manière générale, je me sens à l'aise avec les nombres

- ◯ Pas du tout d'accord
- ◯ Pas d'accord
- ◯ Moyennement d'accord
- ◯ D'accord
- ◯ Tout à fait d'accord

9. Lisez-vous les informations à propos des chances de gains (l'arrière du ticket) ?

- ◯ Oui
- ◯ Non

10. Si oui, vous les trouvez (plusieurs réponses possibles):

- ☐ Suffisamment claires
- ☐ Suffisamment compréhensibles
- ☐ Suffisamment détaillées
- ☐ Pas suffisamment: claires, compréhensibles et détaillées

11. Pensez-vous gagner un jour le gros lot ?

- ◯ Oui
- ◯ Non

12. Selon vous, quelles sont les chances de **récupérer au moins votre mise** après avoir joué **1 fois** ? (Ne répondez que pour les tickets à gratter auxquels vous jouez)

	Très peu probable (0%-20%)	Peu probable (20%-40%)	Probable (40%-60%)	Fort probable (60%-80%)	Très fort probable (80%-100%)
21	○	○	○	○	○
ASTRO	○	○	○	○	○
BINGO	○	○	○	○	○
BLING BLING	○	○	○	○	○
Candy Scratch	○	○	○	○	○
CASH 1€	○	○	○	○	○
CASH 3€	○	○	○	○	○
CASH 5€	○	○	○	○	○
CASH 10€	○	○	○	○	○
KING OF CASH	○	○	○	○	○
MOTS CROISÉS	○	○	○	○	○
PRESTO	○	○	○	○	○
PRESTO EXPERT	○	○	○	○	○
PRESTO UPGRADE	○	○	○	○	○
QUICK CASH	○	○	○	○	○
QUICK CASH 10	○	○	○	○	○
Subito!	○	○	○	○	○
Subito! XL	○	○	○	○	○
Subito! XXL	○	○	○	○	○
VIP CLUB	○	○	○	○	○
WIN FOR LIFE MINI €500	○	○	○	○	○
WIN FOR LIFE	○	○	○	○	○
WIN FOR LIFE DELUXE €3.000	○	○	○	○	○
Winners Club	○	○	○	○	○

13. Selon vous, quelles sont les chances de **récupérer au moins votre mise** après avoir joué **20 fois** ? (Ne répondez que pour les tickets à gratter auxquels vous jouez)

	Très peu probable (0%-20%)	Peu probable (20%-40%)	Probable (40%-60%)	Fort probable (60%-80%)	Très fort probable (80%-100%)
21					
ASTRO					
BINGO					
BLING BLING					
Candy Scratch					
CASH					
CASH					
CASH 5€					
CASH 10€					
KING OF CASH					
MOTS CROISÉS					
PRESTO					
PRESTO EXPERT					
PRESTO UPGRADE					
QUICK CASH					
QUICK CASH					
Subito!					
Subito! XL					
Subito! XXL					
VIP CLUB					
WIN FOR LIFE MINI €500					
WIN FOR LIFE					
WIN FOR LIFE DELUXE €3.000					
Winners Club					

14. Selon vous, quelles sont les chances de gagner **plus que le prix du ticket** après avoir joué **1 fois** ? (Ne répondez que pour les tickets à gratter auxquels vous jouez)

	Très peu probable (0%-20%)	Peu probable (20%-40%)	Probable (40%-60%)	Fort probable (60%-80%)	Très fort probable (80%-100%)
21	○	○	○	○	○
ASTRO	○	○	○	○	○
BINGO	○	○	○	○	○
BLING BLING	○	○	○	○	○
Candy Scratch	○	○	○	○	○
CASH 1€	○	○	○	○	○
CASH 3€	○	○	○	○	○
CASH 5€	○	○	○	○	○
CASH 10€	○	○	○	○	○
KING OF CASH	○	○	○	○	○
MOTS CROISES	○	○	○	○	○
PRESTO	○	○	○	○	○
PRESTO EXPERT	○	○	○	○	○
PRESTO UPGRADE	○	○	○	○	○
QUICK CASH	○	○	○	○	○
QUICK CASH 50	○	○	○	○	○
Subito!	○	○	○	○	○
Subito! XL	○	○	○	○	○
Subito! XXL	○	○	○	○	○
VIP CLUB	○	○	○	○	○
WIN FOR LIFE MINI €500	○	○	○	○	○
WIN FOR LIFE	○	○	○	○	○
WIN FOR LIFE DELUXE €3.000	○	○	○	○	○
Winners Club	○	○	○	○	○

15. Selon vous, quelles sont les chances de gagner **plus que le prix du ticket** après avoir joué **20 fois** ? (Ne répondez que pour les tickets à gratter auxquels vous jouez)

	Très peu probable (0%-20%)	Peu probable (20%-40%)	Probable (40%-60%)	Fort probable (60%-80%)	Très fort probable (80%-100%)
21	○	○	○	○	○
ASTRO	○	○	○	○	○
BINGO	○	○	○	○	○
BLING BLING	○	○	○	○	○
Candy Scratch	○	○	○	○	○
CASH 1€	○	○	○	○	○
CASH 3€	○	○	○	○	○
CASH 5€	○	○	○	○	○
CASH 10€	○	○	○	○	○
KING OF CASH	○	○	○	○	○
MOTS CROISÉS	○	○	○	○	○
PRESTO	○	○	○	○	○
PRESTO EXPERT	○	○	○	○	○
PRESTO UPGRADE	○	○	○	○	○
QUICK CASH	○	○	○	○	○
QUICK CASH	○	○	○	○	○
Subito!	○	○	○	○	○
Subito! XL	○	○	○	○	○
Subito! XXL	○	○	○	○	○
VIP CLUB	○	○	○	○	○
WIN FOR LIFE MINI €500	○	○	○	○	○
WIN FOR LIFE	○	○	○	○	○
WIN FOR LIFE DELUXE €3.000	○	○	○	○	○
Winners Club	○	○	○	○	○

Loterie Nationale

Questions complémentaires

16. Est-ce que vous parlez parfois de la Loterie Nationale (au bureau, à la maison, entre amis, etc.), par exemple de ce que vous feriez si vous gagnez le gros lot ?

- Oui
- Non

17. Est-ce que des personnes de votre entourage jouent aux jeux de la Loterie Nationale (compagne, parent(s), ami(s) proche(s) et/ou enfant(s)) ?

- Oui
- Non

18. Est-ce que l'un de vos parents (ou les deux) jouait régulièrement aux jeux de la Loterie Nationale durant votre enfance ?

- Oui
- Non
- Aucune idée

19. Selon vous, quel pourcentage de ses revenus la Loterie Nationale reverse-t-elle en subsides /sponsoring (à des organisations, projets et événements sportifs, culturels, sociaux, écologiques, scientifiques, etc.) ?

- 80%-100%
- 60%-80%
- 40%-60%
- 20%-40%
- 0%-20%
- Je ne savais pas que la Loterie Nationale reversait une partie de ses revenus en subsides/sponsoring

Loterie Nationale

Profil sociodémographique

20. Quel est votre sexe ?

- ◯ Homme
- ◯ Femme

21. Dans quel tranche d'âge vous situez-vous?

- ◯ moins de 15 ans
- ◯ 15 à 18 ans
- ◯ 19 à 25 ans
- ◯ 26 à 35 ans
- ◯ 36 à 45 ans
- ◯ 46 à 55 ans
- ◯ 56 à 65 ans
- ◯ 66 ou plus

22. Dans quelle ville/commune vivez-vous actuellement ? (Veuillez indiquer le code postale de votre commune)

23. Quel statut décrit le mieux votre situation ?

- ◯ Célibataire
- ◯ En couple
- ◯ Marié(e)
- ◯ Divorcé(e)/séparé(e)
- ◯ Veuf/veuve

24. Avez-vous des enfants à charge ?

- ◯ oui
- ◯ non

25. Quel est votre revenu mensuel (**BRUT**)? *Ceci inclut les revenus professionnels ; les revenus provenant de la location de biens ; toutes pensions, rentes alimentaires et revenus de remplacement ; tous les dividendes et intérêts. Veuillez mentionner le montant en Euros (sans déduire le montant payé en impôts ou toute éventuelle déduction fiscale mentionnée dans votre déclaration d'impôt sur le revenu).*

- ◯ 0 – 999
- ◯ 1 000 - 1 999
- ◯ 2 000 - 2 999
- ◯ 3 000 - 3 999
- ◯ 4 000 - 6 999
- ◯ 8 000 - 9 999
- ◯ 10 000 ou plus
- ◯ Je ne désire pas répondre

26. Parmi les catégories suivantes, laquelle décrit le mieux votre statut professionnel actuel ?

- ◯ Étudiant(e)
- ◯ Employé(e) ou ouvrier/ouvrière - temps partiel
- ◯ Employé(e) ou ouvrier/ouvrière - temps plein
- ◯ Indépendant(e)
- ◯ Profession libérale
- ◯ Sans emploi - recherche un emploi
- ◯ Sans emploi - ne recherche pas d'emploi
- ◯ Pensionné(e)
- ◯ En incapacité de travail
- ◯ Autre (veuillez préciser)

27. Quel est votre niveau d'étude ?

- Sans diplôme
- Enseignement primaire
- 1ère humanité
- 2ème humanité
- 3ème humanité
- 4ème humanité
- 5ème humanité
- Diplômé des études secondaires
- 1ère année d'études supérieur
- 2ème année d'étude supérieur
- Diplômé de bachelier
- 4éme année d'étude supérieur
- Diplômé de master
- Doctorat
- Autre (veuillez préciser)

28. Êtes-vous locataire ou propriétaire de votre maison/appartement ?

- Locataire
- Propriétaire
- Autre (veuillez spécifier)

Annexe 3 : Matrice hypothèse – Items (questions)

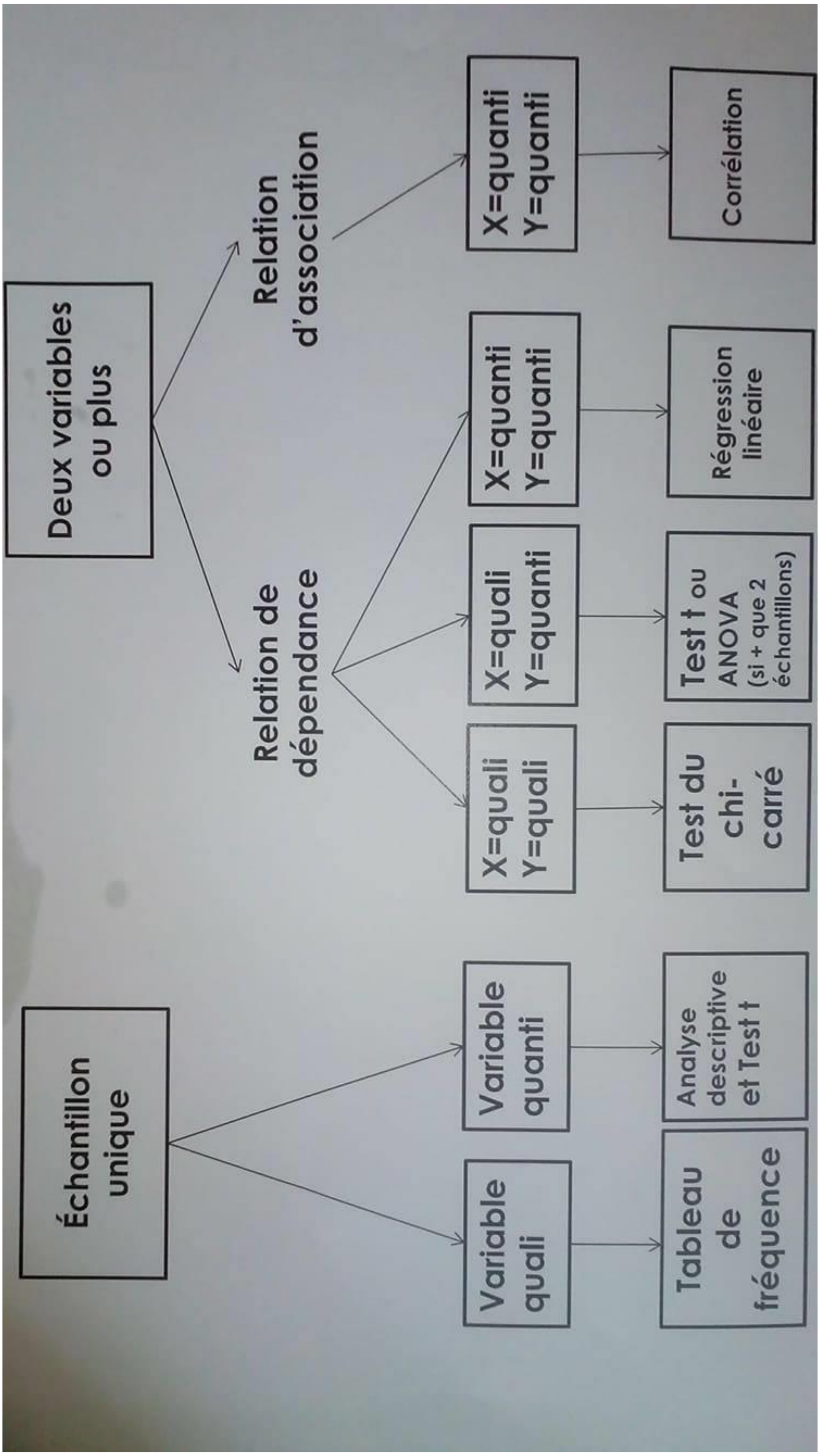
Questions/Hypothèses	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16	H17	H18	H19	H20	H21
Question 1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Question 2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Question 3		X					X				X										
Question 4			X		X	X	X	X	X	X											
Question 5			X		X	X	X	X	X	X											
Question 6												X									
Question 7												X									
Question 8																					
Question 9													X								
Question 10																					
Question 11																X					
Question 12														X	X						
Question 13														X	X						
Question 14														X							
Question 15														X							
Question 16																	X				
Question 17																		X			
Question 18																			X		
Question 19																				X	X
Question 20	X	X	X																		
Question 21				X	X																
Question 22																					
Question 23						X															
Question 24							X														
Question 25								X		X											
Question 26									X												
Question 27										X											
Question 28											X										

Annexe 4 : Questions du sondage triées selon les échelles primaires de mesure

Tableau reprenant les questions triées selon leurs échelles primaires de mesure (nominale, ordinale, d'intervalle ou de proportion) :

Echelles primaires de mesure	Nominale	Ordinale	D'intervalle	De proportion
Question 1	X			
Question 2	X			
Question 3				X
Question 4				X
Question 5				X
Question 6	X			
Question 7	X			
Question 8			X	
Question 9	X			
Question 10	X			
Question 11	X			
Question 12				X
Question 13				X
Question 14				X
Question 15				X
Question 16	X			
Question 17	X			
Question 18	X			
Question 19				X
Question 20	X			
Question 21				X
Question 22	X			
Question 23	X			
Question 24	X			
Question 25				X
Question 26	X			
Question 27		X		
Question 28	X			

Annexe 5 : Fiche récapitulative sur les tests statistiques sélectionnés selon la nature des variables étudiées



Annexe 6 : Hypothèse 1, le taux de participation est plus élevé parmi les hommes.

Nous avons une variable expliquée et une variable explicative.

Variable dépendante : taux de participation aux jeux de grattage (qualitatif nominal)

Variable indépendante : sexe (qualitatif nominal)

H0 : le genre n'a pas d'influence sur le taux de participation aux jeux de grattage.

H1 : le genre a une influence sur le taux de participation aux jeux de grattage

Nous effectuons un test khi-2 : Y : taux de participation aux jeux de grattage (joue 2) et X : sexe

Conditions d'application : pour appliquer le test, il est nécessaire que :

1. Chaque observation soit indépendante des autres (par exemple, elles ne peuvent pas avoir été récoltées chez un même sujet) →respectée

2. Le nombre de cellules ("cases") ayant des fréquences attendues inférieures à 5 soit inférieur à 1/3 du nombre de cellules totales →respectée (voir plus loin dans l'analyse). Chemin sur SPSS : Analyse > Descriptive Statistics > Crosstabs > Row : Sexe et Column : Joue 2 > Cells, cocher observed et expected > Statistics, cocher Chi-square et Phi et Cramers V respectée. H_0:X=0 etH_1:X≠0

Le coefficient Pearson du test khi-carré a une p-valeur de 0,186 (voir tableaux en annexe XXX) ce qui est supérieur à l'alpha (0,05) et le test est donc non significatif et il y a non rejet de H_0. En d'autres mots, les variables « Taux de participation aux jeux de grattage » (joue2) et « sexe » sont indépendantes. Il est donc inutile d'aller plus loin dans l'analyse de ce test khi carré.

Tableau croisé Sexe * Joue2

			Joue2		
			Oui	Non	Total
Sexe	Homme	Effectif	115	25	140
		Effectif théorique	110,7	29,3	140,0
		% dans Sexe	82,1%	17,9%	100,0%

		standard	4,3	-4,3	
	Femme	Effectif	89	29	118
		Effectif théorique	93,3	24,7	118,0
		% dans Sexe	75,4%	24,6%	100,0%
		standard	-4,3	4,3	
Total		Effectif	204	54	258
		Effectif théorique	204,0	54,0	258,0
		% dans Sexe	79,1%	20,9%	100,0%

Tests du khi-deux

	Valeur	ddl	Signification asymptotique (bilatérale)	Sig. exacte (bilatérale)	Sig. exacte (unilatérale)
khi-deux de Pearson	1,747[a]	1	,186		
Correction pour continuité[b]	1,364	1	,243		
Rapport de vraisemblance	1,741	1	,187		
Test exact de Fisher				,220	,121
Association linéaire par linéaire	1,740	1	,187		
N d'observations valides	258				

a. 0 cellule (0,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 24,70.

b. Calculée uniquement pour une table 2x2

Annexe 7 : Hypothèse 2, les hommes ont une fréquence de jeu plus élevée que celle des femmes

Nous avons une variable expliquée et une variable explicative.

Variable dépendante : fréquence de jeu (qualitatif nominal)

Variable indépendante : sexe (qualitatif nominal)

H0 : le genre n'a pas d'influence sur la fréquence de jeu aux jeux de grattage.

H1 : le genre a une influence sur la fréquence de jeu aux jeux de grattage

Nous effectuons un test khi-2 :

Y : fréquence de jeu aux jeux de grattage et X : sexe

Conditions d'application : pour appliquer le test, il est nécessaire que :

1. Chaque observation soit indépendante des autres (par exemple, elles ne peuvent pas avoir été récoltées chez un même sujet) → respectée

2. Le nombre de cellules ("cases") ayant des fréquences attendues inférieures à 5 soit inférieur à 1/3 du nombre de cellules totales → respectée (voir plus loin dans l'analyse). Chemin sur SPSS : Analyse > Descriptive Statistics > Crosstabs > Row : Sexe et Column : Fréquence_de_jeu > Cells, coher observed et expected > Statistics, cocher Chi-square et Phi et Cramers V respectée. H_0:X=0 etH_1:X≠0

Le coefficient Pearson du test Chi-carré a une p-valeur de 0,253 ce qui est supérieure à l'alpha (0,05) et le test est donc non significatif et il y a non rejet de H_0. En d'autres mots, les variables « fréquence de jeu » et « sexe » sont indépendantes. Il est donc inutile d'aller plus loin dans l'analyse de ce test khi carré.

Tableau croisé Sexe * Fréquence_de_jeu

			Fréquence_de_jeu					
			2-3 fois par semaine	1 fois par semaine	1 fois tous les 15 jours	1 fois par mois	moins d'une fois par mois	Total
Sexe	Homme	Effectif	8	28	9	33	37	115
		Effectif théorique	6,8	29,9	11,3	27,1	40,0	115,0
		% dans Sexe	7,0%	24,3%	7,8%	28,7%	32,2%	100,0%
		standard	1,2	-1,9	-2,3	5,9	-3,0	
	Femme	Effectif	4	25	11	15	34	89
		Effectif théorique	5,2	23,1	8,7	20,9	31,0	89,0
		% dans Sexe	4,5%	28,1%	12,4%	16,9%	38,2%	100,0%
		standard	-1,2	1,9	2,3	-5,9	3,0	
Total		Effectif	12	53	20	48	71	204
		Effectif théorique	12,0	53,0	20,0	48,0	71,0	204,0
		% dans Sexe	5,9%	26,0%	9,8%	23,5%	34,8%	100,0%

Tests du khi-deux

	Valeur	ddl	Signification asymptotique (bilatérale)

khi-deux de Pearson	5,353[a]	4	,253
Rapport de vraisemblance	5,451	4	,244
Association linéaire par linéaire	,005	1	,942
N d'observations valides	204		

a. 0 cellules (0,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 5,24.

Annexe 8 : Hypothèse 3, les hommes jouent des sommes plus élevées que les femmes.

Nous avons une variable expliquée et une variable explicative.

Variable dépendante : dépenses mensuelles en jeux de grattage (quantitatif de proportion)

Variable indépendante : sexe (qualitatif nominal)

H0 : en Brabant wallon, le genre n'a pas d'influence sur les dépenses mensuelles en jeux de grattage.

H1 : en Brabant wallon, le genre a une influence sur les dépenses mensuelles en jeux de grattage.

Nous effectuons donc un test t.

H0 : Y (Dépense_mensuelle) = $\beta_0 + erreur$ ou $\beta_2(sexe) = 0$

H1 : Y (Dépense_mensuelle) = $\beta_0 + \beta_2 X_2(sexe)$ ou $\beta_2(sexe) \neq 0$.

Les conditions d'application du test sont les suivantes :

1. Echantillons simples, aléatoires et indépendants.
2. Les scores de la variable scale sont distribués normalement autour de la moyenne dans chaque groupe : test de Kolgomorov Smirnov. Chemin sur SPSS : Analyse > Nonparametric Tests > Legacy Dialogs > 1-Sample K-S Test > insérer Dépense_mensuelle dans Test Variable List > cocher Normal. Cependant la condition n'est pas remplie, dès lors nous avons fait le test de Mann-Whitney. Nous utilisons ce test pour comparer deux groupes (hommes et femmes) par rapport à une variable, à savoir les dépenses mensuelles en jeux de grattage. Chemin sur SPSS : Analyse > Nonparametric Tests > Legacy Dialogs > Two Independent Samples Tests > insérer la variable sexe dans grouping variable et le définir > insérer la variable Dépense_mensuelle dans Test Variable List > cocher Mann-Whitney test.

Test Kolgomorov-Smirnov :

$H0$: La variable scale dépendante se distribue normalement dans les sous-groupes

$H1$: La variable ne se distribue pas normalement

Chemin sur SPSS : *Analyse* > *Nonparametric Tests* > *Legacy Dialogs* > *1-Sample K-S Test* > insérer « Dépense_mensuelle » dans *Test Variable List* > cocher *Normal*.

Test Kolmogorov-Smirnov pour un échantillon

		Dépense_mensuelle
N		204
Paramètres normaux[a,b]	Moyenne	2,75
	Ecart type	1,987
Différences les plus extrêmes	Absolue	,251
	Positif	,251
	Négatif	-,189
Statistiques de test		,251
Sig. asymptotique (bilatérale)		,000[c]

a. La distribution du test est Normale.

b. Calculée à partir des données.

c. Correction de signification de Lilliefors.

Z(dl) = 0,251 ; p = 0,000 ≤ 0,05 donc Rejet H0

Test de Mann-Whitney

Chemin sur SPSS : *Analyze* > *Nonparametric Tests* > *Legacy Dialogs* > *Two Independent Samples Tests* > insérer la variable Sexe dans *Grouping variable* et le définir > insérer la variable Dépense_mensuelle dans *Test Variable List* > cocher *Mann-Whitney test*.

Rangs

	Sexe	N	Rang moyen :	Somme des rangs
Dépense_mensuelle	Homme	115	104,35	12000,00
	Femme	89	100,11	8910,00

Total	204		

Tests statistiques[a]

	Dépense_mensuelle
U de Mann-Whitney	4905,000
W de Wilcoxon	8910,000
Z	-,528
Sig. asymptotique (bilatérale)	,598

a. Variable de regroupement : Sexe

Les résultats de ce test sont les suivants : U = 4905, p-valeur = 0,598 > 0,05 donc non-rejet $H0$. Nous ne pouvons donc pas affirmer que les moyennes des deux groupes sont significativement différentes et par conséquent que le sexe a une influence significative sur les dépenses mensuelles en jeux de grattage.

Annexe 9 : Hypothèse 5, les personnes âgées (61 ans et plus) jouent des sommes plus élevées que les autres.

Nous avons une variable expliquée et une variable explicative.

Variable dépendante : dépenses mensuelles en jeux de grattage (quantitatif de proportion)

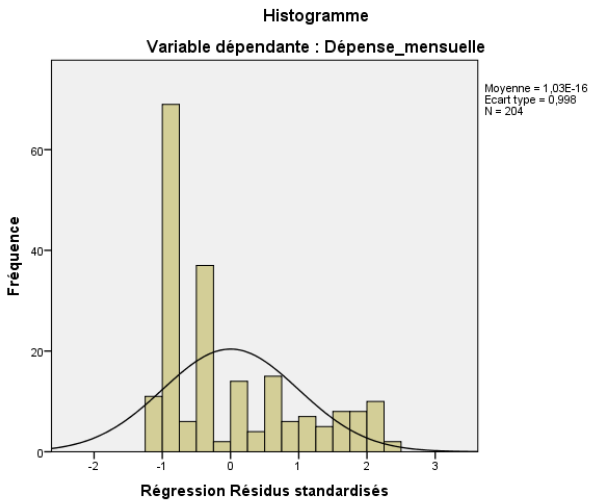
Variable indépendante : âge (quantitatif de proportion)

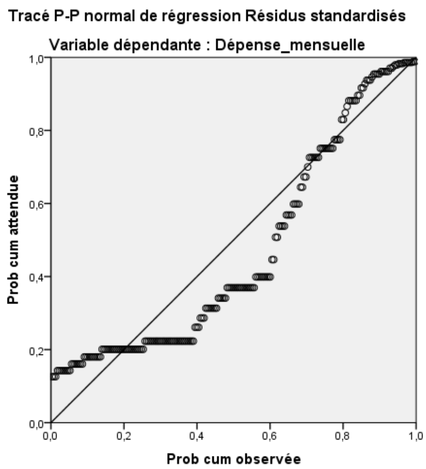
H0 : l'âge a une influence sur les dépenses mensuelles en jeux de grattage.

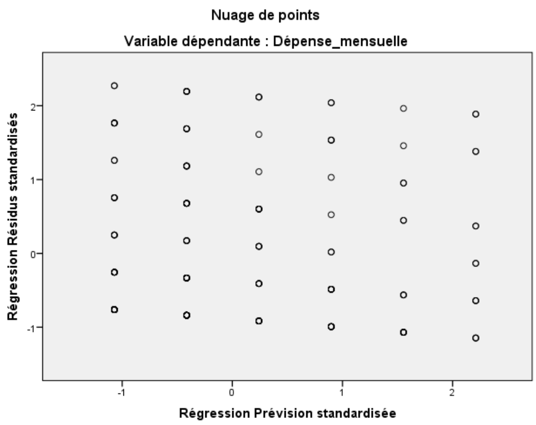
H1 : l'âge n'a pas d'influence sur les dépenses mensuelles en jeux de grattage.

Nous effectuons donc une régression linéaire.

Pour vérifier les hypothèses sous-jacentes et effectuer les régressions nous réalisons : Analyze=>Regression=>Linear=>Sélectionner la variable dépendente et indépendente => Plots => Y : ZRESID et X : ZPRED =>cocher « Histogram » et « Normal probability plot ».







Hypothèses sous-jacentes :

1) histogramme avec résidus de distribution normale moyenne très proche de zéro, tableau 3.21 (non valide).

2) P-P plot de distribution normale car données proches de la droite de Henri, tableau 3.22 (non valide).

3) Scatterplot vérifie qu'il n'y a pas d'hétéroscédasticité donc que les données sont bien réparties autour de zéro (valide) et que la distribution est normale (95% des valeurs entre -2 et +2), tableau 3.23 (non valide) et que le comportement des résidus semble aléatoire (nous sommes en présence d'une variable issue de plusieurs catégories de réponses, il est donc normal que les résidus soient disposés de cette manière).

Puisqu'aucune des hypothèses sous-jacentes n'est respectée, nous ne pouvons pas effectuer de régression linéaire.

Annexe 10 : Hypothèse 6, les joueurs mariés ont des dépenses mensuelles plus élevées que les joueurs célibataires.

Nous avons une variable expliquée et une variable explicative.

Variable dépendante : dépenses mensuelles en jeux de grattage (quantitatif de proportion)

Variable indépendante : état civil (qualitatif nominal, comprenant plus de 2 catégories)

H0 : la situation conjugale a une influence sur les dépenses mensuelles jeux de grattage.

H1 : la situation conjugale n'a pas d'influence sur les dépenses mensuelles en jeux de grattage.

Nous effectuons donc un test ANOVA

H0 : Y (Dépense_mensuelle) = $\beta_0 + erreur$ ou β_2($s$ituation_conjugale) = 0

H1 : Y (Dépense_mensuelle) = $\beta_0 + \beta_2 X_2$(situation_conjugale) ou β_2(situation_conjugale) ≠ 0.

Pour réaliser les ANOVAS et les tests de Welch : Analyze=>Compare means=> One-Way ANOVA=>sélectionner les variables (dépendante et indépendante) =>Options=> cocher « Homogeneity of variance test » (ANOVA) ou « Welch » (pour test de Welch).

Test ANOVA :

Test d'homogénéité des variances

Dépense_mensuelle

Statistique de Levene	ddl1	ddl2	Sig.
2,934	4	199	,022

ANOVA

Dépense_mensuelle

	Somme des carrés	ddl	Carré moyen	F	Sig.
Inter-groupes	47,127	4	11,782	3,107	,017
Intragroupes	754,618	199	3,792		
Total	801,745	203			

Homogénéité des variances : dans le test de Levene la p-valeur doit être supérieure à 0.05, ce qui n'est pas le cas ici, donc nous ne pouvons pas affirmer que les variances sont équivalentes dans la population. Il faut donc réaliser un test de Welch.

Test de Welch :

Tests robustes d'égalité des moyennes

Dépense_mensuelle

	Statistiques[a]	ddl1	ddl2	Sig.
Welch	3,043	4	35,164	,030

a. F distribué asymptotiquement

La p-valeur associée au test F qui est de 0,03 est inférieure à 0,05 nous pouvons donc affirmer avec un risque d'erreur de 5% que dans la population **les dépenses mensuelles diffèrent significativement selon la situation conjugale**.

En analysant notre base de données, nous avons ainsi remarqué que les dépenses mensuelles s'élèvent à plus de 20€ par mois pour :

- 10,25% des joueurs célibataires

- 31,25% des joueurs divorcés
- 20,83% des joueurs en couple
- 22,23% des joueurs mariés

Pour ce qui est des veufs, 0% joue plus de 20€ par mois (sur un échantillon assez faible de 8 individus, et donc probablement peu représentatif de la réalité).

Nous constatons donc effectivement des différences, par exemple il semblerait que ce soit les joueurs divorcés qui jouent des sommes plus élevées que les autres joueurs.
L'hypothèse est quant à elle vérifiée, puisque les mariés sont deux fois plus nombreux que les célibataires à jouer 20€ et plus par mois.

Annexe 11 : Hypothèse 7, les joueurs ayant un/des enfants à charge dépensent moins d'argent dans les jeux à gratter

Nous avons une variable expliquée et une variable explicative.

Variable dépendante : dépenses mensuelles en jeux de grattage (quantitatif de proportion)

Variable indépendante : enfant(s) à charge (qualitatif nominal)

H0 : Avoir des enfants à charge a une influence sur les dépenses mensuelles en jeux de grattage.

H1 : Avoir des enfants à charge n'a pas d'influence sur les dépenses mensuelles en jeux de grattage.

Nous faisons donc un test t.

H0 : Y (Dépense_mensuelle) = $\beta_0 + erreur$ ou β_2(Enfants) = 0

H1 : Y (Dépense_mensuelle) = $\beta_0 + \beta_2\, X_2(sexe)$ ou β_2(Enfants) $\neq$ 0.

Les conditions d'application du test sont les suivantes :

1. Les échantillons doivent être simples, aléatoires et indépendants.

2. Les scores de la variable scale sont distribués normalement autour de la moyenne dans chaque groupe: test de Kolgomorov Smirnov. Chemin sur SPSS : Analyse > Nonparametric Tests > Legacy Dialogs > 1-Sample K-S Test > insérer dépense_mensuelle dans Test Variable List > cocher Normal. Cependant la condition n'est pas remplie, dès lors nous avons fait le test de Mann-Whitney. Nous utilisons ce test pour comparer deux groupes (hommes et femmes) par rapport à une variable, à savoir le nombre d'heures d'utilisation d'Internet. Chemin sur SPSS : Analyse > Nonparametric Tests > Legacy Dialogs > Two Independent Samples Tests > insérer la variable Enfants dans grouping variable et le définir > insérer la variable dépense_mensuelle dans Test Variable List > cocher Mann-Whitney test.

Test Kolgomorov-Smirnov :

$H0$: La variable scale dépendante se distribue normalement dans les sous-groupes

$H1$: La variable ne se distribue pas normalement

Chemin sur SPSS : *Analyse* > *Nonparametric Tests* > *Legacy Dialogs* > *1-Sample K-S Test* > insérer « Dépense_mensuelle » dans *Test Variable List* > cocher *Normal*.

Test Kolmogorov-Smirnov pour un échantillon

		Dépense_mensuelle
N		204
Paramètres normaux[a,b]	Moyenne	2,75
	Ecart type	1,987
Différences les plus extrêmes	Absolue	,251
	Positif	,251
	Négatif	-,189
Statistiques de test		,251
Sig. asymptotique (bilatérale)		,000[c]

a. La distribution du test est Normale.

b. Calculée à partir des données.

c. Correction de signification de Lilliefors.

Z(dl) = 0,251 ; p = 0,000 ≤ 0,05 donc Rejet H0

Test de Mann-Whitney

Chemin sur SPSS : *Analyze* > *Nonparametric Tests* > *Legacy Dialogs* > *Two Independent Samples Tests* > insérer la variable Enfants dans *Grouping variable* et le définir > insérer la variable Dépense_mensuelle dans *Test Variable List* > cocher *Mann-Whitney test*.

Rangs

	Enfants	N	Rang moyen :	Somme des rangs
Dépense_mensuelle	Oui	62	108,44	6723,00
	Non	142	99,91	14187,00

Total	204		

Tests statistiques[a]

	Dépense_mensuelle
U de Mann-Whitney	4034,000
W de Wilcoxon	14187,000
Z	-,985
Sig. asymptotique (bilatérale)	,324

a. Variable de regroupement : Enfants

Les résultats de ce test sont les suivants : U = 4034, p-valeur = 0,324 > 0,05 donc non-rejet $H0$. Nous ne pouvons donc pas affirmer que les moyennes des deux groupes sont significativement différentes et par conséquent que le fait d'avoir des enfants à charge a une influence significative sur les dépenses mensuelles en jeux de grattage.

Annexe 12 : Hypothèse 8, les individus ayant les niveaux de revenu les plus bas tendent à allouer une plus grande proportion de leur revenu dans les tickets à gratter.

Nous avons une variable expliquée et une variable explicative.

Variable dépendante : dépenses mensuelles en jeux de grattage (quantitatif de proportion)

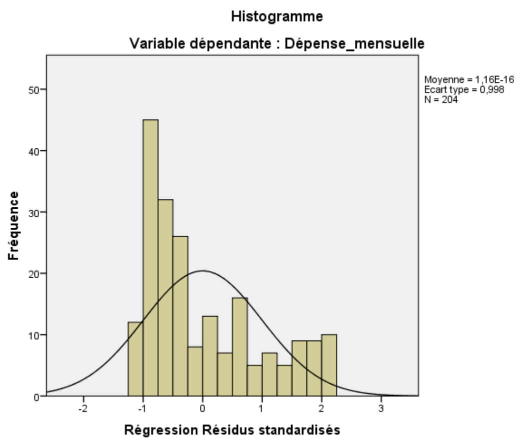
Variable indépendante : niveaux de revenus (quantitatif de proportion)

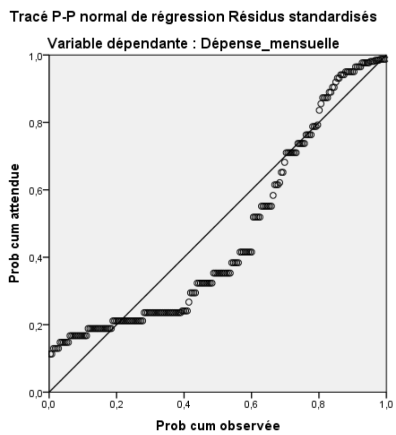
H0 : le niveau de revenu a une influence sur les dépenses mensuelles en jeux de grattage.

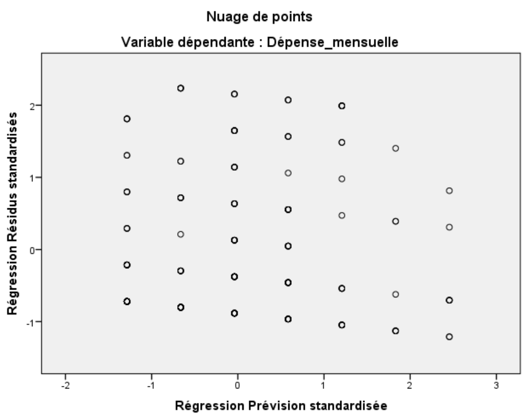
H1 : le niveau de revenu n'a pas d'influence sur les dépenses mensuelles en jeux de grattage.

Nous réalisons donc une régression linéaire.

Pour vérifier les hypothèses sous-jacentes et effectuer les régressions nous réalisons : Analyze=>Regression=>Linear=>Sélectionner la variable dépendante et indépendante=>Plots=>Y : ZRESID et X : ZPRED =>cocher « Histogram » et « Normal probability plot ».







Hypothèses sous-jacentes :

1. histogramme avec résidus de distribution normale moyenne très proche de zéro (non valide)
2. P-P plot de distribution normale car données proches de la droite de Henri (non valide)
3. Scatterplot vérifie qu'il n'y a pas d'hétéroscédasticité donc que les données sont bien réparties autour de zéro (valide) et que la distribution est normale (95% des valeurs entre -2 et +2) (non valide) et que le comportement des résidus semble aléatoire (nous sommes en présence d'une variable issue de plusieurs catégories de réponses, il est donc normal que les résidus soient disposés de cette manière)

Puisqu'aucune des hypothèses sous-jacentes n'est respectée, nous ne pouvons pas faire de régression linéaire.

Annexe 13 : Hypothèse 9, les joueurs possédant le statut d'employé ont des dépenses mensuelles plus élevées que ceux ayant d'autres statuts d'emploi (indépendant, étudiant, retraité, sans-emploi).

Nous avons une variable expliquée et une variable explicative.

Variable dépendante : dépenses mensuelles en jeux de grattage (quantitatif de proportion)

Variable indépendante : statut professionnel (qualitatif nominal)

H0 : le statut professionnel a une influence sur les dépenses mensuelles en jeux de grattage.

H1 : le statut professionnel n'a pas d'influence sur les dépenses mensuelles en jeux de grattage.

Nous effectuons donc un test ANOVA

H0 : Y (Dépense_mensuelle) = $\beta_0 + erreur$ ou β_2(Statut_professionnel) = 0

H1 : Y (Dépense_mensuelle) = $\beta_0 + \beta_2 X_2$(Statut_professionnel) ou β_2(Statut_professionnel) $\neq 0$.

Pour effectuer les ANOVAS et les tests de Welch : Analyze=>Compare means=> One-Way ANOVA=>sélectionner les variables (dépendante et indépendante) =>Options=>cocher « Homogeneity of variance test » (ANOVA) ou « Welch » (pour test de Welch).

Test ANOVA :

Test d'homogénéité des variances

Dépense_mensuelle

Statistique de Levene	ddl1	ddl2	Sig.
,714	7	196	,660

Homogénéité des variances : dans le test de Levene la p-valeur doit être supérieure à 0.05, ce qui est le cas ici, nous pouvons donc affirmer que les variances sont équivalentes dans la population.

ANOVA

Dépense_mensuelle

	Somme des carrés	ddl	Carré moyen	F	Sig.
Inter-groupes	17,452	7	2,493	,623	,737
Intragroupes	784,293	196	4,001		
Total	801,745	203			

La p-valeur associée au test F est supérieure à 0.05 nous ne pouvons donc pas affirmer avec un risque d'erreur de 5% que dans la population, les dépenses mensuelles en jeux de grattage diffèrent significativement selon le statut professionnel. Il y a donc rejet de H0.

Annexe 14 : Hypothèse 10, les joueurs moins instruits dépensent plus en tickets à gratter.

Nous avons une variable expliquée et une variable explicative.

Variable dépendante : dépenses mensuelles en jeux de grattage (quantitatif de proportion)

Variable indépendante : niveau d'éducation (qualitatif nominal)

H0 : le niveau d'éducation a une influence sur les dépenses mensuelles jeux de grattage.

H1 : le niveau d'éducation n'a pas d'influence sur les dépenses mensuelles en jeux de grattage.

Nous effectuons donc un test ANOVA

H0 : Y (Dépense_mensuelle) = $\beta_0 + erreur$ ou β_2(Niveau_d_étude) = 0

H1 : Y (Dépense_mensuelle) = $\beta_0 + \beta_2 X_2$(Niveau_d_étude) ou β_2(Niveau_d_étude) $\neq$0.

Pour effectuer les ANOVAS et les tests de Welch : Analyze=>Compare means=> One-Way ANOVA=>sélectionner les variables (dépendante et indépendante) =>Options=>cocher « Homogeneity of variance test » (ANOVA) ou « Welch » (pour test de Welch).

Test ANOVA :

Test d'homogénéité des variances

Dépense_mensuelle

Statistique de Levene	ddl1	ddl2	Sig.
1,817	7	196	,086

Homogénéité des variances : dans le test de Levene la p-valeur doit être supérieure à 0.05, ce qui est le cas ici, nous pouvons donc affirmer que les variances sont équivalentes dans la population.

ANOVA

Dépense_mensuelle

	Somme des carrés	ddl	Carré moyen	F	Sig.
Inter-groupes	33,144	7	4,735	1,207	,300

Intragroupes	768,601	196	3,921		
Total	801,745	203			

La p-valeur associée au test F est supérieure à 0.05. Nous ne pouvons donc pas affirmer avec un risque d'erreur de 5% que dans la population, les dépenses mensuelles en jeux de grattage diffèrent significativement selon le statut professionnel. Il y a donc rejet de H0.

Annexe 15 : Hypothèse 11, les individus propriétaires jouent moins fréquemment que les individus locataires.

Nous avons une variable expliquée et une variable explicative.

Variable dépendante : fréquence de jeux (qualitatif nominal)

Variable indépendante : être propriétaire (qualitatif nominal)

H0 : le fait d'être propriétaire a une influence sur les dépenses mensuelles en jeux de grattage.

H1 : le fait d'être propriétaire n'a pas d'influence sur les dépenses mensuelles en jeux de grattage.

Nous effectuons un test khi-2 :

Y : fréquence de jeu aux jeux de grattage et X : Logement

Conditions d'application : pour appliquer le test, il est nécessaire que :

1. Chaque observation soit indépendante des autres (par exemple, elles ne peuvent pas avoir été récoltées chez un même sujet) → respectée

2. Le nombre de cellules ("cases") ayant des fréquences attendues inférieures à 5 soit inférieures à 1/3 du nombre de cellules totales → respectée (voir plus loin dans l'analyse). Chemin sur SPSS : Analyse > Descriptive Statistics > Crosstabs > Row : Logement et Column : Fréquence_de_jeu > Cells, coher observed et expected > Statistics, cocher Chi-square et Phi et Cramers V respectée. H_0:X=0 etH_1:X≠0

Le coefficient Pearson du test Chi-carré a une p-valeur de 0,405 ce qui est supérieur à l'alpha (0,05) et le test est donc non significatif et il y a non rejet de H_0. En d'autres mots, les variables « Fréquence_de_jeu » et « Logement » sont indépendantes. Il est donc inutile de poursuivre dans l'analyse de ce test khi carré.

Tableau croisé Logement * Fréquence_de_jeu

Effectif

		Fréquence_de_jeu					
		2-3 fois par semaine	1 fois par semaine	1 fois tous les 15 jours	1 fois par mois	moins d'une fois par mois	Total
Logement	Locataire	8	30	8	22	35	103
	Propriétaire	4	19	11	18	25	77
	Vit chez ses parents	0	4	1	8	11	24
Total		12	53	20	48	71	204

Tests du khi-deux

	Valeur	ddl	Signification asymptotique (bilatérale)
khi-deux de Pearson	8,298[a]	8	,405
Rapport de vraisemblance	9,632	8	,292
Association linéaire par linéaire	3,209	1	,073
N d'observations valides	204		

a. 3 cellules (20,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 1,41.

Annexe 16 : Hypothèse 17 : Les individus qui en parlent avec leur entourage jouent plus.

Nous avons une variable expliquée et une variable explicative.

Variable dépendante : le taux de participation aux jeux de grattage (qualitatif nominal)

Variable indépendante : le fait d'en parler avec l'entourage participation (qualitatif nominal)

H0 : le fait d'en parler avec l'entourage a une influence sur le taux de participation aux jeux de grattage.

H1 : le fait d'en parler avec l'entourage n'a pas d'influence sur le taux de participation aux jeux de grattage.

Nous effectuons un test khi-2 :

Y : Joue2 et X : Sujet_de_discussion

Conditions d'application : pour appliquer le test, il est nécessaire que :

1. Chaque observation soit indépendante des autres (par exemples, elles ne peuvent pas avoir été récoltées chez un même sujet) → respectée

2. Le nombre de cellules ("cases") ayant des fréquences attendues inférieures à 5 soit inférieur à 1/3 du nombre de cellules totales → respectée (voir plus loin dans l'analyse). Chemin sur SPSS : Analyse > Descriptive Statistics > Crosstabs > Row : Sujet_de_discussion et Column : Joue2 > Cells, coher observed et expected > Statistics, cocher Chi-square et Phi et Cramers V respectée. $H0{:}X{=}0$ et $H1{:}X{\neq}0$

Tableau croisé Sujet_de_discussion * Joue2

			Joue2		
			Oui	Non	Total
Sujet_de_discussion	Oui	Effectif	173	34	207
		Effectif théorique	163,7	43,3	207,0
	Non	Effectif	31	20	51
		Effectif théorique	40,3	10,7	51,0
Total		Effectif	204	54	258

Effectif théorique	204,0	54,0	258,0

Sur la ligne « Expected Count », c'est ce qu'on devrait observer dans notre échantillon s'il n'y avait aucun lien entre les deux variables. La ligne « Count » est ce qu'on observe réellement dans l'échantillon. Ceci permet de voir de manière intuitive s'il y a un lien ou non entre les deux variables. Pour voir si ces différences sont significatives, il faut aller voir le coefficient beta du test chi-carré (ci-dessous).

Tests du khi-deux

	Valeur	ddl	Signification asymptotique (bilatérale)	Sig. exacte (bilatérale)	Sig. exacte (unilatérale)
khi-deux de Pearson	12,842[a]	1	,000		
Correction pour continuité[b]	11,502	1	,001		
Rapport de vraisemblance	11,500	1	,001		
Test exact de Fisher				,001	,001
Association linéaire par linéaire	12,793	1	,000		
N d'observations valides	258				

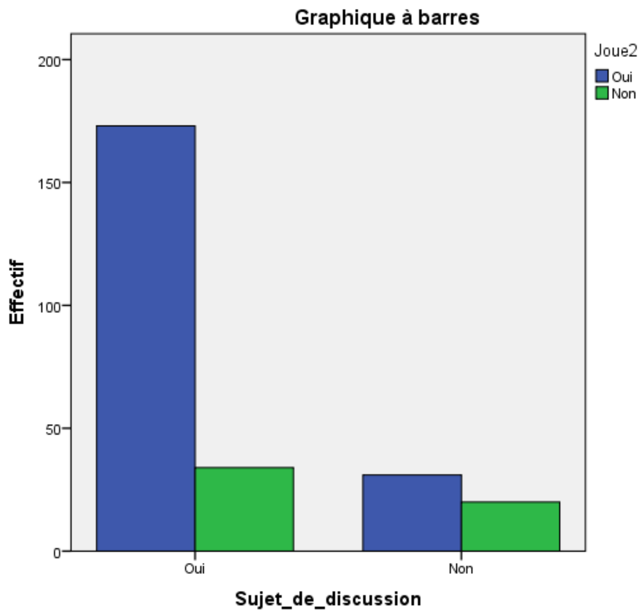
a. 0 cellules (0,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 10,67.

b. Calculée uniquement pour une table 2x2

La sous hypothèse est respectée étant donné que nous avons moins de 20% dans la parenthèse de la note de bas de tableau « a. 0 cells (0,0%) have expected count less than 5. »

Le coefficient Pearson du test Chi-carré a une p-valeur de 0,001 ce qui est inférieur à l'alpha (0,05) et le test est donc significatif et il y a rejet de $H0$. En d'autres mots, il semble que les variables « Sujet_de_discussion » et « Joue2 » ont une relation de dépendance.

Effectivement, au vu du graphique ci-dessous, nous pouvons aisément identifier cette corrélation entre le fait de parler avec son entourage des jeux de loterie et le fait de jouer à ces derniers. Ceci va donc dans le sens de notre hypothèse initiale, à savoir que le fait d'en parler avec l'entourage a une influence sur le taux de participation aux jeux de grattage. Il est néanmoins possible que la relation soit inverse, càd que c'est le fait de participer aux jeux de loterie qui pousse les individus à en parler avec leur entourage.


Graphique à barres
Joue2
Oui
Non
200
150
100
50
0
Effectif
Oui
Non
Sujet_de_discussion

Annexe 17 : Hypothèse 18, les individus dont l'entourage joue, ont tendance à jouer également.

Nous avons une variable expliquée et une variable explicative.

Variable dépendante : le taux de participation aux jeux de grattage en jeux de grattage (qualitatif nominal)

Variable indépendante : Le fait que l'entourage joue (qualitatif nominal)

H0 : le fait que l'entourage joue a une influence sur le taux de participation aux jeux de grattage.

H1 : le fait que l'entourage joue n'a pas d'influence sur le taux de participation aux jeux de grattage.

Nous faisons un test khi-2 :

Y : Joue2 et X : Entourage_jouent

Conditions d'application : pour appliquer le test, il est nécessaire que :

1. Chaque observation soit indépendante des autres (par exemple, elles ne peuvent pas avoir été récoltées chez un même sujet) → respectée

2. Le nombre de cellules ("cases") ayant des fréquences attendues inférieures à 5 soit inférieur à 1/3 du nombre de cellules totales → respectée (voir plus loin dans l'analyse). Chemin sur SPSS : Analyse > Descriptive Statistics > Crosstabs > Row : Entourage_jouent et Column : Joue2 > Cells, coher observed et expected > Statistics, cocher Chi-square et Phi et Cramers V respectée. H_0:X=0 etH_1:X≠0

Tableau croisé Entourage_jouent * Joue2

			Joue2		
			Oui	Non	Total
Entourage_jouent	Oui	Effectif	186	37	223
		Effectif théorique	176,3	46,7	223,0

	Non	Effectif	18	17	35
		Effectif théorique	27,7	7,3	35,0
Total		Effectif	204	54	258
		Effectif théorique	204,0	54,0	258,0

Sur la ligne « Expected Count », c'est ce qu'on devrait observer dans notre échantillon s'il n'y avait aucun lien entre les deux variables. La ligne « Count » est ce qu'on observe réellement dans l'échantillon. Ceci permet de voir de manière intuitive s'il y a un lien ou non entre les deux variables. Pour voir si ces différences sont significatives, il faut aller voir le coefficient beta du test chi-carré (ci-dessous).

Tests du khi-deux

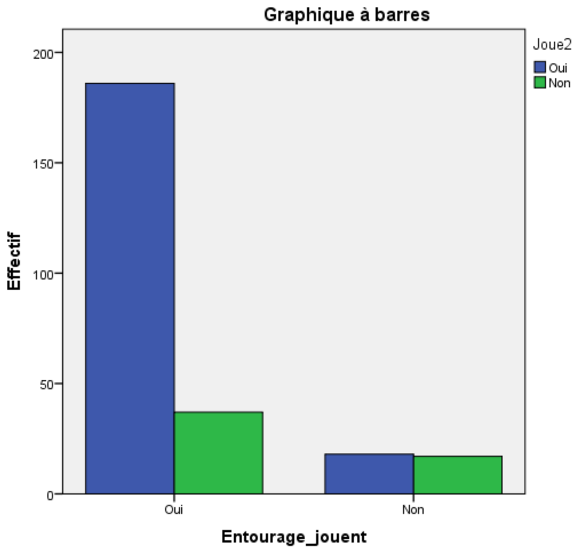
	Valeur	ddl	Signification asymptotique (bilatérale)	Sig. exacte (bilatérale)	Sig. exacte (unilatérale)
khi-deux de Pearson	18,694[a]	1	,000		
Correction pour continuité[b]	16,812	1	,000		
Rapport de vraisemblance	15,819	1	,000		
Test exact de Fisher				,000	,000
Association linéaire par linéaire	18,622	1	,000		
N d'observations valides	258				

a. 0 cellules (0,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 7,33.

b. Calculée uniquement pour une table 2x2

La sous-hypothèse est respectée étant donné que nous avons moins de 20% dans la parenthèse de la note de bas de tableau « a. 0 cells (0,0%) have expected count less than 5. »

Le coefficient Pearson du test Chi-carré a une p-valeur de 0,000 ce qui est inférieur à l'alpha (0,05) et le test est donc significatif et il y a rejet de $H0$. En d'autres mots, il semble que les variables « Entourage_jouent » et « Joue2 » ont une relation de dépendance.



Il semblerait donc que notre hypothèse soit vérifiée et le fait que l'entourage joue a une influence sur le taux de participation aux jeux de grattage. Le graphique ci-dessus semble nous indiquer que les individus dont l'entourage joue ont tendance à eux-mêmes participer aux jeux de loterie.

Annexe 18 : Hypothèse 19, les individus dont les parents jouaient durant leur enfance, ont tendance à jouer également.

Nous avons une variable expliquée et une variable explicative.

Variable dépendante : le taux de participation aux jeux de grattage (qualitatif nominal)

Variable indépendante : le fait que les parents jouaient durant leur enfance (qualitatif nominal)

H0 : le fait que les parents jouaient durant leur enfance a une influence sur le taux de participation aux jeux de grattage.

H1 : le fait que les parents jouaient durant leur enfance n'a pas d'influence sur le taux de participation aux jeux de grattage.

Nous effectuons un test khi-2 :

Y : Joue2 et X : Parents_jouaient

Conditions d'application : pour appliquer le test, il est nécessaire que :

1. Chaque observation soit indépendante des autres (par exemple, elles ne peuvent pas avoir été récoltées chez un même sujet) → respectée

2. Le nombre de cellules ("cases") ayant des fréquences attendues inférieures à 5 soit inférieur à 1/3 du nombre de cellules totales → respectée (voir plus loin dans l'analyse). Chemin sur SPSS : Analyse > Descriptive Statistics > Crosstabs > Row : Parents_jouaient et Column : Joue2 > Cells, coher observed et expected > Statistics, cocher Chi-square et Phi et Cramers V□respectée. $H_0 : X=0$ et $H_1 : X \neq 0$

Tableau croisé Parents_jouaient * Joue2

			Joue2		
			Oui	Non	Total
Parents_jouaient	Oui	Effectif	116	17	133
		Effectif théorique	105,2	27,8	133,0
	Non	Effectif	60	32	92

		Effectif théorique	72,7	19,3	92,0
	Aucune idée	Effectif	28	5	33
		Effectif théorique	26,1	6,9	33,0
Total		Effectif	204	54	258
		Effectif théorique	204,0	54,0	258,0

Sur la ligne « Expected Count », c'est ce qu'on devrait observer dans notre échantillon s'il n'y avait aucun lien entre les deux variables. La ligne « Count » est ce qu'on observe réellement dans l'échantillon. Ceci permet de voir de manière intuitive s'il y a un lien ou non entre les deux variables. Pour voir si ces différences sont significatives, il faut aller voir le coefficient beta du test chi-carré (ci-dessous).

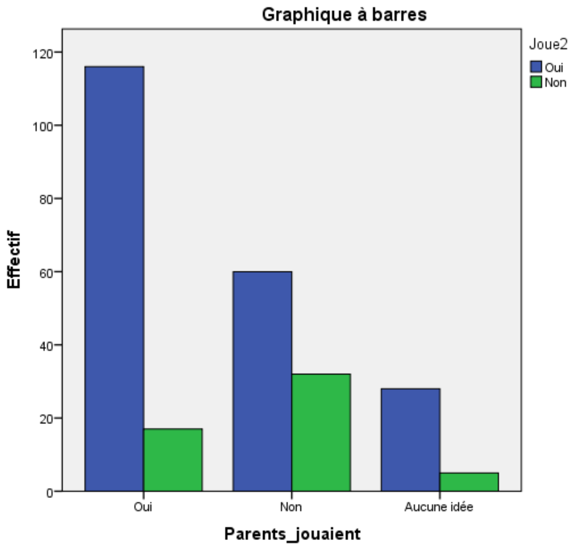
Tests du khi-deux

	Valeur	ddl	Signification asymptotique (bilatérale)
khi-deux de Pearson	16,669[a]	2	,000
Rapport de vraisemblance	16,101	2	,000
Association linéaire par linéaire	3,773	1	,052
N d'observations valides	258		

a. 0 cellules (0,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 6,91.

La sous-hypothèse est respectée étant donné que nous avons moins de 20% dans la parenthèse de la note de bas de tableau « a. 0 cells (0,0%) have expected count less than 5. »

Le coefficient Pearson du test Chi-carré a une p-valeur de 0,000 ce qui est inférieur à l'alpha (0,05) et le test est donc significatif et il y a rejet de $H0$. En d'autres mots, il semble que les variables « Parents_jouaient » et « Joue2 » ont une relation de dépendance.



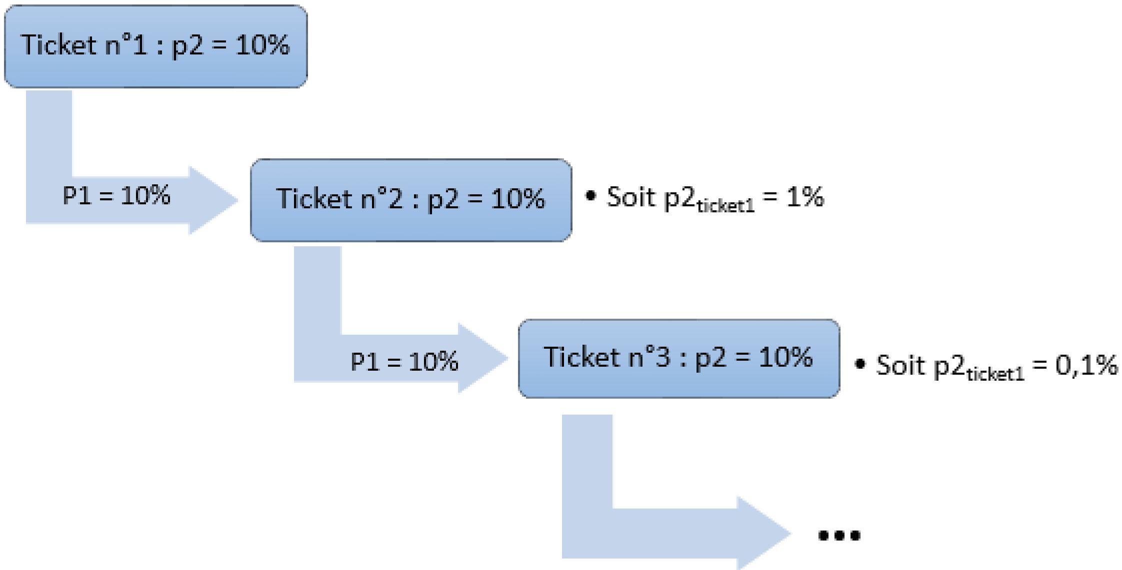
Tout comme pour les deux hypthèse précédentes, il semble que l'hypothèse 19 soit vérifiée. En effet, nous avons une relation de dépendance entre les variables « Parents_jouaient » et « Joue2 » nous laissant croire que le fait que lorsque les parents d'un individu participaient activement aux jeux de loterie durant son enfance, celui-ci a tendance à également jouer par après.

Annexe 19 : Récapitulatif des tests statistiques réalisés

N° hypothèse	Intitulé de l'hypothèse	Type de test effectué	p-valeur	Résultat
1	Le taux de participation est plus élevé parmi les hommes	Test khi-carré	0,186	Non rejet de H0
2	Les hommes ont une fréquence de jeu plus élevée que celle des femmes	Test khi-carré	0,253	Non rejet de H0
3	Les hommes jouent des sommes plus élevées que les femmes	Test t	0,598	Non rejet de H0
5	Les personnes âgées (61 ans et plus) jouent des sommes plus élevées que les autres	Régression linéaire	/	Hypothèses sous-jacentes non respectées
6	Les joueurs mariés ont des dépenses mensuelles plus élevées que les joueurs célibataires	Test ANOVA	0,03	Rejet H0 : les dépenses mensuelles diffèrent significativement selon la situation conjugale
7	Les joueurs ayant un/des enfant(s) à charge dépensent moins d'argent dans les jeux à gratter	Test t	0,324	Non rejet de H0
8	Les individus ayant les niveaux de revenu les plus bas tendent à allouer une plus grande proportion de leur revenu dans les tickets à gratter	Régression linéaire		Hypothèses sous-jacentes non respectées
9	Les joueurs possédant le statut d'employé ont des dépenses mensuelles plus élevées que ceux ayant d'autres statuts d'emploi (indépendant, étudiant, retraité, sans-emploi)	Test ANOVA	0,737	Non rejet de H0
10	Les joueurs moins instruits dépensent plus en tickets à gratter	Test ANOVA	0,3	Non rejet de H0
11	Les individus propriétaires jouent moins fréquemment que les individus locataires	Test khi-carré	0,405	Non rejet de H0

17	Les individus qui en parlent avec leur entourage jouent plus	Test khi-carré	0,001	Rejet H0 : les variables « Sujet_de_discussion » et « Joue2 » ont une relation de dépendance. Il y a une corrélation entre le fait de parler avec son entourage des jeux de loterie et le fait de jouer à ces derniers.
18	Les individus dont l'entourage joue, ont tendance à jouer également	Test khi-carré	0	Rejet H0 : les variables « Entourage_joue » et « Joue2 » ont une relation de dépendance. Le fait que l'entourage joue a une influence sur le taux de participation aux jeux de grattage.
19	Les individus dont les parents jouaient durant leur enfance, ont tendance à jouer également	Test khi-carré	0	Rejet H0 : les variables « Parents_jouaient » et « Joue2 » ont une relation de dépendance. Lorsque les parents d'un individu participaient activement aux jeux de loterie durant son enfance, celui-ci a tendance à également jouer par après.

Annexe 20 : la méthodologie suivie pour le ticket « Candy Scratch » et « 21 »



Annexe 21 : Explication méthodologique de la récursion de Panjer.

Ci-dessous se trouvent une explication méthodologique de la récursion de Panjer. Cet extrait provient de l'article « Win or Lose for Life ? Regard croisé sur les jeux de hasard et les produits structurés. » de Vrins et Petitjean (2016).

4.2.2. *Récursion de Panjer*

Une solution alternative, qui ne passe pas par le domaine fréquentiel, se base sur la récursion de Panjer (1981). Nous procédons itérativement et construisons la distribution de $Y^{k+1} = Y^k + X_{k+1} = \sum_{i=1}^{k+1} X_i$ en convoluant la distribution de $Y^k = \sum_{i=1}^{k} X_i$ avec celle de X_{k+1}, c'est-à-dire celle de X dans le cas où les variables additionnées sont indépendantes et identiquement distribuées. Nous supposerons que X peut prendre les valeurs $\{X_0, \ldots, X_m\}$.

Le point de départ ($k = 1$) correspond évidemment à imposer $Y^1 = X_1$, dont la distribution p_X est donnée. Détaillons l'étape d'induction, qui consiste à obtenir la distribution de Y^{k+1}

à partir de celle de Y^k qui est supposée connue. Nous noterons $1+j_k$ le nombre de valeurs possibles pour cette variable, de sorte que la distribution de Y^k est donnée par l'ensemble des couples (y_j^k, q_j^k) pour $j = \{0,1,2,\ldots,j_k\}$. On se rappellera que $\sum_{j=0}^{j_k} q_j^k = 1$. Partant de chacune des valeurs atteignables y_j^k de Y^k, on peut déduire les valeurs possibles pour Y^{k+1} obtenues suite à l'ajout de la variable X_{k+1}. Pour $j = \{0,1,2,\ldots,j_k\}$ et $i =\{0,1,\ldots,m\}$, on obtient donc le gain $y_{i,j,k} = y_j^k + x_i$. Cette situation correspond au cas $\left(Y^k = y_j^k, X_{k+1} = x_i\right)$ dont la probabilité est obtenue par indépendance :

$$q_{i,j,k} = P(Y^k = y_j^k, X_{k+1} = x_i) = P(Y^k = y_j^k)\, P(X_{k+1} = x_i) = q_j^k\, q_i.$$

En toute généralité, il existe plusieurs couples (j,k) permettant d'atteindre une même valeur pour $y_{i,j,k}$. On regroupera donc les séquences menant à des valeurs identiques pour $y_j^k + x_i$. Nous noterons $1 + j_{k+1}$ le nombre de valeurs distinctes atteignables pour $y_j^k + x_i$ qui peuvent être obtenues au moyen des différentes combinaisons (i,j) de manière à obtenir les valeurs strictement croissantes $0, y_1^{k+1}, \ldots, y_{j_{k+1}}^{k+1}$ que peut prendre la variable Y^{k+1}. Pour $j = \{0,1,2,\ldots j_{k+1}\}$, nous noterons l_j^k la liste des couples (i,j), tels que $y_{i,j,k} = y_j^{k+1}$, de sorte que

$$y_{i,j,k} = y_j^{k+1} \textit{ si et seulement si } (i,j) \in \mathrm{l}_j^k.$$

La liste l_j^k nous donne donc les différentes manières d'obtenir une valeur donnée pour Y^{k+1} en additionnant les variables indépendantes Y^k et X_{k+1} . On obtient donc la probabilité associée à y_j^{k+1} en additionnant les probabilités liées aux différentes combinaisons possibles :

$$q_j^{k+1} = \sum_{i,j} 1_{\{y_{i,j,k}=y_j^{k+1}\}} q_{i,j,k} = \sum_{(i,j)\in l_j^k} q_{i,j,k}.$$

Nous obtenons enfin les probabilités $q_1^{k+1}, \dots, q_{j_{k+1}}^{k+1}$ associées aux valeurs $y_1^{k+1}, \dots, y_{j_{k+1}}^{k+1}$ qui correspond à la distribution de probabilité de Y^{k+1}.

Illustrons le fonctionnement de la procédure de récursion de Panjer sur l'exemple du « Win for Life », avec $n = 2$. La distribution de $G^1 = G_1$ est connue : il s'agit de la distribution des gains obtenus en une participation, qui est donnée dans le Tableau 1. Elle est définie comme l'ensemble des couples $\{(x_i,p_i)\}$ où $i = \{0,1,2,\dots,m\}$ avec $m = 10$. Ces couples nous donnent l'ensembles des valeurs possibles pour G^1, que nous noterons $\{(g_i,q_i)\}$ dans notre procédure de récursion. La première valeur possible pour G^2 est $y_0^2 = 0$. On trouvera qu'il n'existe qu'une seule manière d'obtenir cette valeur, c'est d'avoir $y_{0,0,k}$, c'est-à-dire (G^1,G_2)=(0,0), et donc l_0^1 $=\{(0,0)\}$. Dès lors, la probabilité associée est donnée par $q_0^2 = q_0^1 q_0 = q_0\, q_0$. La seconde plus grande valeur possible est $y_1^2 = 3$. Il existe deux couples (i,j) permettant d'aboutir à cette valeur, car $y_{0,1,k} = y_{1,0,k} = 3$. En effet, on obtient un gain total de 3 euros en deux participations en gagnant une seule fois. On trouve donc la liste $l_1^1 = \{(0,1),(1,0)\}$. On trouver la probabilité associée à y_1^2 en sommant les probabilité $q_{i,j,1}$ sur le domaine $(i,j) \in l_1^1$: $q_1^2 = q_{0,1,1} + q_{1,0,1}$=$q_1^1 q_0 + q_0^1 q_1 = 2q_0\, q_1$. De manière similaire, on trouvera que la troisième valeur possible est $y_2^2 = 6$. Il est évident que l'on peut obtenir un gain cumulé de 6 de trois manières différentes pour (G_1,G_2) : {(3,3), (0,6), (6,0)}. En terme d'indices, on trouvera la liste $\mathrm{l}_2^1 = \{(1,1),(0,2),(2,0)\}$. La probabilité associée est donc la somme des probabilités associées à chacun de ces cas : $q_2^2 = q_{1,1,1} + q_{0,2,1} + q_{2,0,1} = q_1^1 q_1 + q_0^1 q_2 + q_2^1 q_0$. On trouvera qu'il existe j_2 valeurs possibles pour G^2. Cette dernière valeur correspond également au cas unique $\mathrm{l}_{j_2}^1 = \{(j_1, 10)\}$ qui correspond au gain maximum (où l'on gagne à chaque fois) et dont la probabilité est $q_{j_2}^2 = q_{j_1,10,1} = q_{j_1}^1 q_{10}$. On peut continuer à itérer de cette manière pour atteindre la distribution d'une somme de variables. A chaque étape, on doit bien vérifier que

$$\mathrm{q}_j^k > 0 \ \ et \ \sum_{j=0}^{j_k} \mathrm{q}_j^k = 1.$$

Il est important de réaliser que le procédure de Panjer est entièrement automatique : il n'est pas nécessaire d'identifier les différentes manières d'obtenir un certain gain, même si cela peut servir à des fins de vérification et d'explication. En pratique, tout ceci est implicite et encapsulé dans l'algorithme.

Annexe 22 Le code de Panjer implémenté dans le programme R avec la distribution du « Cash €3 ».

Voici le code de Panjer implémenté dans le programme R avec à titre d'exemple la distribution du « Cash €3 ». Ce code nous as été fourni par notre promoteur, M. Vrins.

```r
# distribution de la moyenne des gains de n billets de loterie de gains spécifiés

# source("PanjerGain3.R")

# ==============================================================
# Inputs
# ==============================================================

n = 20 # nb of participations, integer >1
pu = 3 # cost of 1 ticket

# gain distribution in 1 participation

g = c(100000,10000,5000,1000,500,100,50,30,20,15,9,6,3,0) # vector of gains
p = c(2,2,4,4,20,1080,5000,24400,15000,34500,27000,148000,258500)/1000  # vector

# choose method to compute the distribution of gains in n participations

method = "Panjer"   # "Panjer" or Monte Carlo, "MC"

output = "detailed"   # "detailed" or "simple", only used in case of method == Par
M1 = 500000     # nb of simulations, only used in case of method == MC

# ==============================================================
# Pre-calculations
# ==============================================================

ng = 1-sum(p)   # proba that gain = 0
p  = c(p,ng)    # vector of probabilities (incl. gain=0)
np = length(p)    # nb of possible gains

moy = p%*%g-pu    # expected gain (net, ie incl. cost of ticket)
print(c(paste("gain moyen/billet : ",moy)))

# plot distribution (not easy to read because x hugely spread)
```

```r
# plot distribution (not easy to read because x hugely spread)
plot(g-pu,p,col="blue",pch=19)
for(i in 1:np){
  points(rep(g[i]-pi,2),c(0,p[i]),type="l",col="blue",lty=2)
}

p.cum = cumsum(p) # vector of cumulative probabilities
M = n*M1     # only used in case of method == MC

# ============================================================
# Calculations of gain distribution in n particpations
# ============================================================

if(method=="Panjer"){
  # this is an analytical method that delivers the EXACT distribution

  grid.g = g         # grid of attainable gains
  grid.p = p         # associated probabilities
  mu = p%*%g         # expected value (gross)
  sd = sqrt(sum(p*g*g)-mu^2)   # standard deviation

  # << start Panjer recursion >>

  for(i in 2:n){
    # include ticket i

    lg  = length(grid.g)
    pjk = matrix(0,ncol=np,nrow=lg) # pjk is the probabilities to get in this stat
    gjk = pjk             # cumulative gain levels with ticket i knowing that (i) ga

    for(j in 1:lg){
      # add possible gains g[k] to any possible previous levels gj
      pj = grid.p[j]
      gj = grid.g[j]
      for(k in 1:np){
```

5:2 (Untitled) R Script

```
      for(k in 1:np){
        gjk[j,k] = gj+g[k] # add gain g[k] to previous level gj
        pjk[j,k] = pj*p[k] # proba to have a cumulative gain gj (pj) without ticke
      }
    }

    # order gains
    idx = order(gjk)
    grid.g = gjk[idx] # get ordered list of attainable gains
    grid.p = pjk[idx] # associated prob
    # sum(grid.p) # this is just to check that this is equal to 1 (otherwise prob

    # now we have to merge the probabilities associated to same levels of gains --
    lg = length(grid.g)
    grid.g.new = c(grid.g[1])
    grid.p.new = c(grid.p[1])
    for(j in 2:lg){
      if(grid.g[j]>grid.g[j-1]){
        grid.g.new = c(grid.g.new,grid.g[j])
        grid.p.new = c(grid.p.new,grid.p[j])
      }else{
        grid.p.new[length(grid.p.new)] = grid.p.new[length(grid.p.new)] + grid.p[j
      }
    }
    grid.g = grid.g.new
    grid.p = grid.p.new
  }

  # << end Panjer recursion >>

  #grid.p[1] # this is the proba to have total gain=0, which must be equal to (p[r

  P=cumsum(grid.p)  # cumulative gain probabilities
  G=grid.g    # cumulative gains
```

95:2 (Untitled) R Script

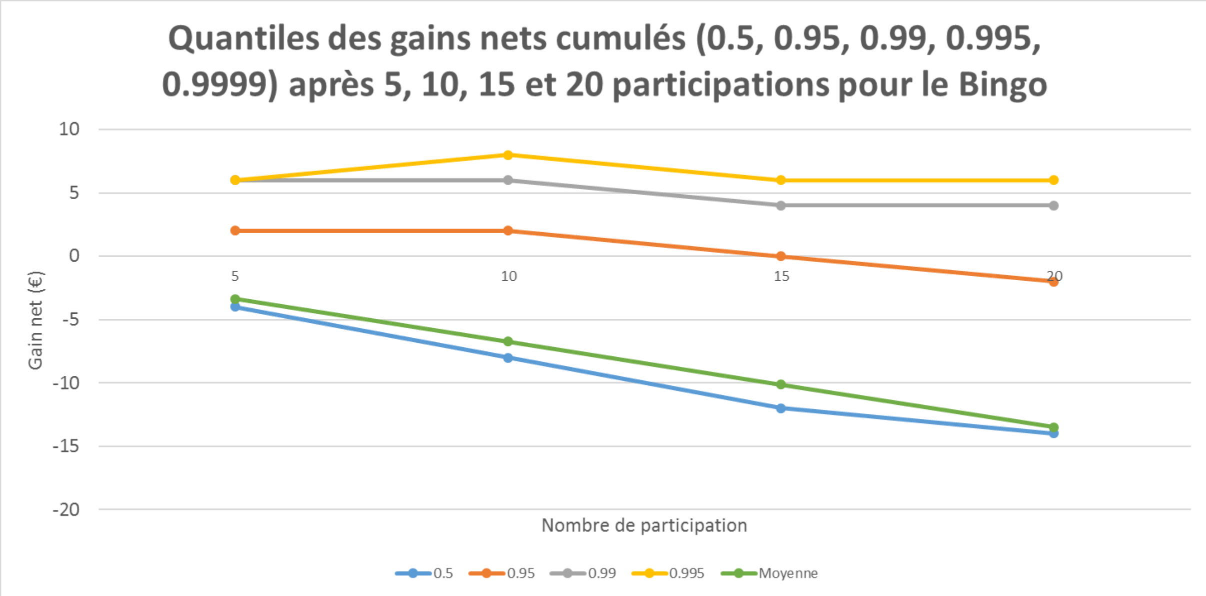
```r
  G=grid.g    # cumulative gains

  # plot resulting distribution (again, not easy to read)
  dev.new()
  plot(G-n*pu,grid.p)

  # some indicators about the total gains' distribution :
  # -------------------------------------------------------

  # proba to get back at least the cost, ie gross gain >= n*pu or equivalently net
  idx = which(grid.g>=n*pu)
  print(c(paste("proba de récupérer ses gains : ",sum(grid.p[idx]))))

  # proba to get in total MORE than cost, ie gross gain > n*pu or equivalently net
  idx = which(grid.g>n*pu)
  print(c(paste("proba de récupérer plus que ses gains : ",sum(grid.p[idx]))))


}else{

  # Monte Carlo : this is a numerical method that delivers an APPROXIMATION of the

  # Generate M gains (M1=M/n for each participation)
  U = runif(M)
  G = rep(0,M)
  for(i in 1:M){
    idx=which(U[i]<p.cum)
    G[i]=g[idx[1]]
  }
  G = matrix(G,nrow=n,ncol=M1)  # Gij represents the gain obtain when playing the
  Gs = colSums(G)-n*pu    # net gain after n parts wrt run

  # get some indicators about the distribution
  p.looser=length(which(Gs<0))/M1
  idx=which(Gs>=0)
```

195:2 (Untitled) R Script

```r
  idx=which(Gs>=0)
  p.winner=1-p.looser
  print(c(paste("proba de récupérer sa mise : ", p.winner)))     # estimation of th
  quantile(Gs[idx],0.90)
  print(c(paste("quantile 90% de ceux qui ont au moins récupérer leur mise : ", me
  idx=which(Gs<0)
  quantile(Gs[idx],0.95)
  print(c(paste("quantile 95% de ceux qui n'ont pas récupérer leur mise : ", mean(
}

if( (method=="Panjer") && (output=="detailed") ){

  grid.ps=cumsum(grid.p)
  pv=c(0.5,0.95,0.99,0.995,0.9999)
  nq=length(pv)
  qv=rep(0,nq)
  egv=qv
  for(i in 1:nq){
    idx=which(grid.ps>pv[i])
    qv[i]=grid.g[idx[1]]
    egv[i]=(grid.g[idx]%*%grid.p[idx])/sum(grid.p[idx])
  }
  qv=qv-n*pu
  egv=egv-n*pu
  print(paste(c("quantiles : ",pv)))
  print(paste(c("quantiles gains nets cumulés : ",qv)))

  # get net positive cumulative gains
  idx=which(grid.g>=n*pu)
  # conditional prob
  pc=(grid.p[idx])/sum(grid.p[idx])
  grid.gi=grid.g[idx]
  print(c("Moyenne nette cumulée des non-perdants : ",grid.gi%*%pc-n*pu))

  i1=idx[1]
```

```r
  i1=idx[1]
  1-grid.ps[i1-1] #proba to get at least one's money back (=1-sum(grid.p[idx]))
  pcs=cumsum(pc)
  idx=which(pcs>0.90)
  i2=idx[1]
  pcs[i2]
  grid.gi[i2]-n*pu# gain net en cas de gain avec max 90%
  print(c("gains cumulés max 90%",grid.gi[i2])) # gain en cas de gain avec max 90%

  idx=seq(1,(i1-1)) # cas menant à ne pas récupérer sa mise
  pc=(grid.p[idx])/sum(grid.p[idx])
  pcs=cumsum(pc)   # proba conditionelle de gain | pas récupérer sa mise
  grid.gi=grid.g[idx]
  grid.gi%*%pc-n*pu # perte moyenne sachant qu'on n'a pas récupéré sa mise
  print(c("pertes cumulées moyenne des non-gagnants",n*pu-grid.gi%*%pc))

  idx=which(pcs>=0.90)
  i3=idx[1]
  print(c("pertes cumulées max 90%",n*pu-grid.gi[i3]))

  (1-grid.ps[i1-1])*(1-pcs[i2]) # proba d'etre dans les 1%

  #gain max à 99.99% en cas de gains nets >=0
  idx=which(pcs>0.9999)
  grid.g[i1+idx[1]-1]
}
```

Annexe 23 : Quantiles des gains nets cumulés (0.5, 0.95, 0.99, 0.995, 0.9999) après 5, 10, 15 et 20 participations pour le Bingo



Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve, Belgique **www.uclouvain.be/lsm**