

# STATISTIQUES DESCRIPTIVES : ETUDE D'UN JEU DE DONNÉES

Christian INGOUFF  
Yann CHARBONNIER  
Vincent ALVES

EISTI

MAINC1 - Groupe de M. Bourhattas

2013/2014  
Semestre 2

# Table des matières

<b>1</b>	<b>Description du projet</b>	<b>2</b>
<b>2</b>	<b>Présentation de l'échantillon</b>	<b>3</b>
2.1	Introduction . . . . .	3
2.2	Description du jeu de données . . . . .	3
2.3	Jeu de données . . . . .	5
<b>3</b>	<b>Etude statistique</b>	<b>6</b>
3.1	Etude univariée . . . . .	6
3.1.1	Variables qualitatives . . . . .	6
3.1.2	Variables quantitatives . . . . .	8
3.2	Etude bivariée . . . . .	10
3.2.1	Qualitatif x qualitatif . . . . .	10
3.2.2	Quantitatif x qualitatif . . . . .	14
3.2.3	Quantitatif x quantitatif . . . . .	15

# Chapitre 1

## Description du projet

Le projet nous amène à étudier un échantillon de taille supérieure à 30 qui constituera la population d'étude. L'étude sera faite selon les méthodes enseignées dans le module de statistiques descriptives enseigné à l'EISTI.

Ce rapport présente les résultats de cette approche appliquée à un jeu de données choisi au préalable, ainsi que leur interprétation concrète effectuée par notre groupe. Le travail effectué expose de cette manière l'utilisation d'outils pour répondre à une problématique se rapportant au jeu de données.

# Chapitre 2

## Présentation de l'échantillon

### 2.1 Introduction

Le jeu de données fourni présente les 50 jeux provenant de *Steam* les plus joués à un moment précis. *Steam* est une plateforme de distribution de jeux vidéo jouables sur ordinateur. Cette plateforme connaît un grand succès et est fréquentée chaque jour par plusieurs centaines de milliers d'utilisateurs dans le monde entier.

Le succès des jeux vidéo est un fait avéré et reconnu. Nous souhaitons étudier dans ce travail les axes de ce succès sur différents jeux ainsi que les paramètres de ce succès. Est-ce que le type économique du jeu a une incidence sur son succès ? Est-ce que les jeux multijoueurs sont les plus populaires ? Quels sont les types de jeux les plus addictifs, par l'étude des heures de jeu ?

### 2.2 Description du jeu de données

Nous citons les sources suivantes à la date exacte du 23 mai 2014 à 12h30 (heure française) :

- Steam Charts : <http://steamcharts.com/>
- Statistiques de Steam : <http://store.steampowered.com/stats/>

Les variables proposées dans ce jeu de données sont les suivantes :

- *Game* : nom du jeu
  - Variable qualitative nominale
  - Représentable en liste
  
- *Current Players* : nombre de joueurs sur un jeu en même temps à l'heure donnée (variable quantitative, représentable en barres)
  - Variable quantitative discrète
  - Représentable en diagramme en bâtons, par jeu
  
- *Type* : type économique d'un jeu
  - Variable qualitative nominale, de modalités :
    - *Free To Play* : jeu gratuit d'accès
    - *Regular Game* : jeu non gratuit d'accès, qui nécessite son achat
    - *Early Access* : jeu en avant-première ou en version beta
  - Représentable en diagramme camembert
  
- *Peak Players* : record du nombre de joueurs en même temps sur un jeu
  - Variable quantitative discrète
  - Représentable en diagramme en bâtons, par jeu
  
- *Total Hours Played* : nombre d'heures de jeu accumulées par l'ensemble des joueurs d'un jeu
  - Variable quantitative continue (cependant arrondie à l'entier)
  - Représentable en diagramme en bâtons, par jeu
  
- *Multiplayer* : désigne si un jeu est multijoueur ou non
  - Variable qualitative dichotomique (oui ou non)
  
  - Représentable en diagramme en bâtons

## 2.3 Jeu de données

Game	Current Players	Type	Peak Players	Total Hours Played	Multi-player ?
Dota 2	547453	Free To Play	843024	329503280	Yes
Counter-Strike : Global Offensive	51229	Regular Game	170137	60091852	Yes
Team Fortress 2	36540	Free To Play	80053	34003037	Yes
Football Manager 2014	23014	Regular Game	67747	24404659	Yes
Sid Meier's Civilization V	18465	Regular Game	57627	22466607	No
The Elder Scrolls V : Skyrim	12847	Regular Game	44749	16596939	No
Counter-Strike	12512	Regular Game	33755	12729635	Yes
Garry's Mod	12469	Regular Game	48674	17438357	Yes
DARK SOULS™ II	11193	Regular Game	79528	22552539	Yes
DayZ	10589	Early Access	36653	12675446	Yes
Rust	10037	Early Access	25328	10164995	Yes
War Thunder	9323	Free To Play	26228	6981577	Yes
Counter-Strike : Source	8473	Regular Game	24626	8903681	Yes
Warframe	7863	Free To Play	20623	8884552	Yes
Arma 3	7733	Regular Game	46089	10308585	Yes
Terraria	7249	Regular Game	26328	6991031	Yes
Total War : ROME II	6300	Regular Game	18867	6887654	Yes
Left 4 Dead 2	5519	Regular Game	14900	6094640	Yes
Wolfenstein : The New Order	4839	Regular Game	11808	552987	No
Tom Clancy's Ghost Recon Phantoms - NA	4280	Free To Play	16496	5601058	Yes
Football Manager 2013	4166	Regular Game	11521	4178388	Yes
Borderlands 2	3414	Regular Game	12649	4731792	Yes
Neverwinter	3254	Free To Play	8220	2813337	Yes
PAYDAY 2	3117	Regular Game	11116	3848722	Yes
Path of Exile	2632	Free To Play	8630	3398269	Yes
Call of Duty : Black Ops II - Multiplayer	2579	Regular Game	11805	3748688	Yes
Space Engineers	2548	Early Access	8285	3158064	Yes
Arma 2 : Operation Arrowhead	2537	Regular Game	8583	3088087	Yes
Europa Universalis IV	2430	Regular Game	7614	2913904	Yes
Call of Duty : MW3 - Multiplayer	2358	Regular Game	8116	2577244	Yes
Mount & Blade : Warband	2212	Regular Game	9006	2866436	No
The Binding of Isaac	2177	Regular Game	5800	1883122	No
Fallout : New Vegas	2056	Regular Game	6737	2607749	No
PlanetSide 2	2035	Free To Play	6713	2470522	Yes
Tom Clancy's Ghost Recon Phantoms - EU	2034	Free To Play	12885	3476213	Yes
Empire : Total War	2031	Regular Game	7840	2471256	No
Chivalry : Medieval Warfare	1955	Regular Game	13351	1652516	Yes
Crusader Kings II	1937	Regular Game	6090	2420256	No
Euro Truck Simulator 2	1930	Regular Game	7423	2154047	No
Company of Heroes 2	1867	Regular Game	12957	2657977	Yes
Football Manager 2012	1839	Regular Game	5243	1827744	Yes
Kerbal Space Program	1790	Early Access	6823	2492169	No
Starbound	1709	Early Access	5405	2094810	Yes
Dark Souls : Prepare to Die Edition	1697	Regular Game	4923	1733197	Yes
Call of Duty : Ghosts - Multiplayer	1686	Regular Game	12745	2575394	Yes
Total War : SHOGUN 2	1640	Regular Game	4508	1810160	Yes
XCOM : Enemy Unknown	1597	Regular Game	4465	1722967	No
Transistor	1534	Regular Game	6657	243591	No
FTL : Faster Than Light	1521	Regular Game	7714	2567369	No
RIFT™	1429	Free To Play	2974	1299327	Yes

FIGURE 2.1 – Top 50 des jeux les plus joués sur Steam au 23 mai 2014 à 12h30

# Chapitre 3

## Etude statistique

### 3.1 Etude univariée

#### 3.1.1 Variables qualitatives

Nous représentons ci-dessous la répartition des modalités de types de jeu et de multijouabilité parmi les éléments de notre échantillon sous la forme de diagrammes circulaires :

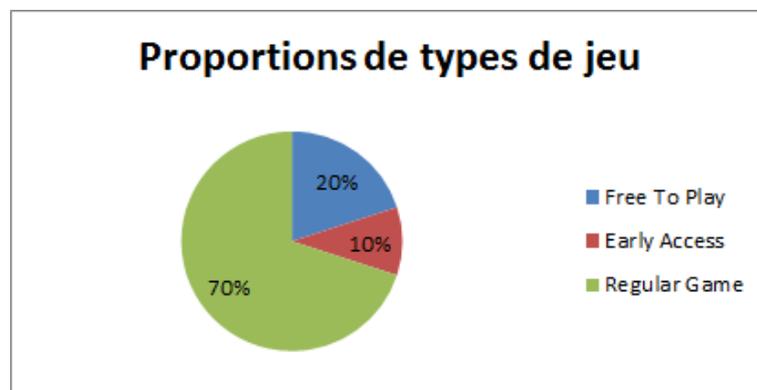


FIGURE 3.1 – Camembert : types de jeu

On constate clairement la prédominance de “jeux normaux”, à savoir le format classique “copie du jeu à acheter”. Cela s’explique aisément par le fait que l’industrie du jeu vidéo s’est bâtie comme beaucoup d’autres sur le modèle capitaliste de la vente d’un bien pour de l’argent en échange.

On ne néglige cependant pas la forme économique davantage moderne du “Free To Play”, qui se traduit par un accès gratuit au jeu (la rentabilité du jeu va se justifier par des achats proposés par le jeu même). En effet, 20% des jeux populaires sur Steam sont des jeux “Free To Play”.

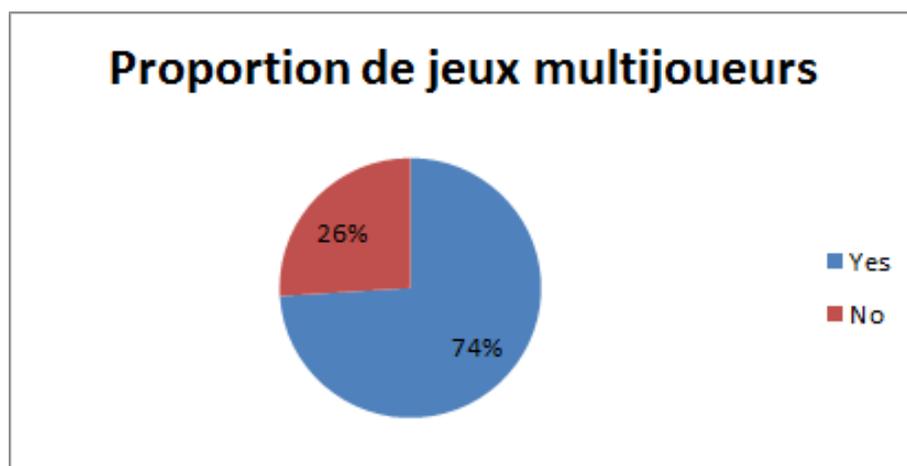


FIGURE 3.2 – Camembert : jeux multijoueurs

Selon ce graphique, il y a environ trois fois plus de jeux multijoueurs que de jeux qui se jouent seul sur la scène populaire de Steam. Ce résultat exprime un intérêt certain des joueurs dans l’interactivité avec d’autres joueurs sur Internet par rapport à un gameplay plus isolé.

Le succès de ce mode de jeu a aussi amené les éditeurs de jeu à implémenter des modes multijoueurs si possible dans leurs jeux. C’est aujourd’hui la norme d’intégrer un mode multijoueur dans des jeux de genres tels que les FPS (First Person Shooters ; jeux de tirs), les jeux de stratégie, les jeux de courses, et ainsi de suite.

### 3.1.2 Variables quantitatives

Nous représentons ci-dessous la répartition des valeurs du nombre de joueurs actuels, du nombre de joueurs record, et du nombre d'heures de jeu sous la forme de boîtes à moustaches (représentées avec une échelle logarithmique) :

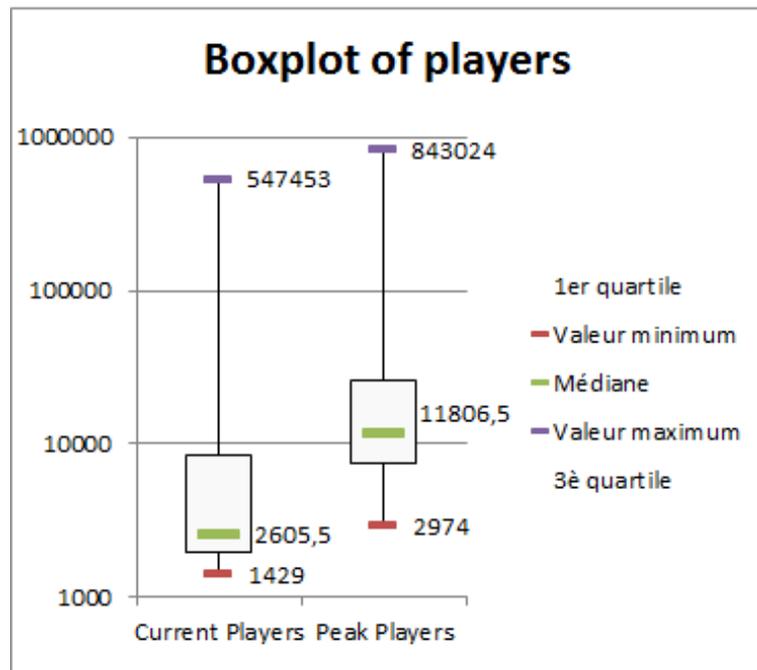


FIGURE 3.3 – Boîte à moustaches : joueurs

Cette première boîte à moustache a dû être modélisée avec une échelle logarithmique à cause des valeurs supérieures très fortement supérieures à la moyenne. En effet, les chiffres du jeu “DotA 2” surpasse de plusieurs fois ceux du deuxième en liste “Counter Strike : Global Offensive”. Cette boîte à moustache montre donc une très forte dispersion supérieure des données.

En comparant le premier quartile de joueurs présents  $Q_1 = 1941,5$  et le troisième quartile  $Q_3 = 8320,5$ , nous pouvons confirmer la forte disparité des chiffres.

On explique le résultat par le phénomène de “jeux à la mode”. En effet, “DotA 2” est un jeu appartenant à la catégorie des “MOBA” (Multiplayer Online Battle Arena ; autrement dit, environnement de combat multijoueur en ligne), genre de jeu étant excessivement populaire (“DotA 2” est par ailleurs largement surpassé par le jeu du même genre “League of Legends”, qui n’est pas représenté par la plateforme Steam).

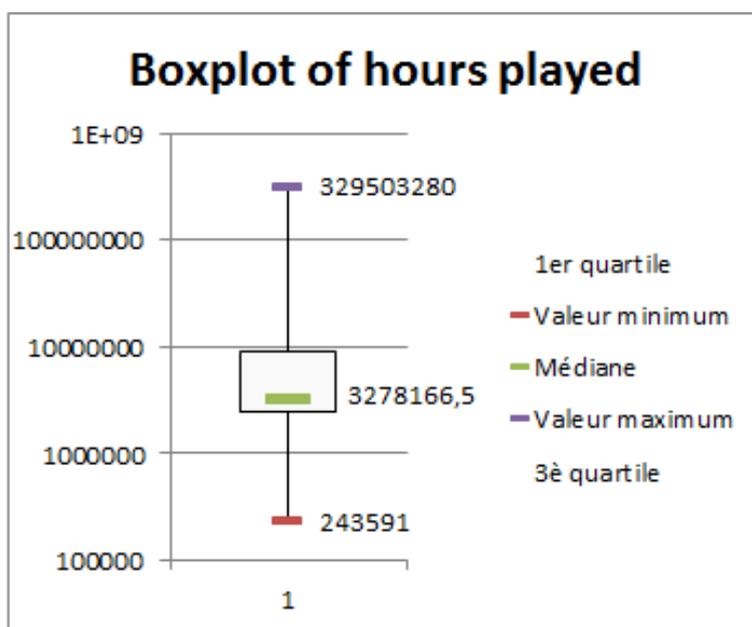


FIGURE 3.4 – Boîte à moustaches : heures de jeu

Cette boîte à moustache a également été représentée avec une échelle logarithmique pour des raisons similaires à la boîte précédente. Ici, nous observons une dispersion des deux côtés : supérieure avec les grands jeux populaires, mais aussi inférieure.

La dispersion supérieure est également expliquée par le phénomène de mode. Mais à cela s’ajoute le paramètre de pérennité du jeu : est-ce que le jeu vaut le coup d’être joué encore et encore ? Le jeu “DotA 2” remplit cette condition, le jeu s’apparentant à un jeu à la fois d’action et de stratégie en temps réel, intégrant ainsi une rejouabilité à la manière d’un jeu d’échecs.

De l'autre côté du spectre se trouvent soit les jeux récents ou les jeux à scénario. Les jeux récents n'ont eu de toute évidence pas le temps d'accumuler beaucoup d'heures de jeu. Quant aux jeux à scénario, ce sont des jeux qui mènent le joueur à travers une histoire, et une histoire ne peut se lire à trop de reprises. La durée de vie de ces jeux ont par conséquent pour moyenne entre 15 et 50h par joueur.

## 3.2 Etude bivariée

### 3.2.1 Qualitatif x qualitatif

Type	Free to Play	Early Access	Regular Game	Sommes
Multijoueur				
Oui	10	4	23	37
Non	0	1	12	13
Sommes	10	5	35	50

FIGURE 3.5 – Multijouabilité par rapport au type de jeu

Nous présentons ici l'étude de la multijouabilité d'un jeu par rapport à son type économique. Le tableau ci-dessus expose le nombre de jeux appartenant à chaque croisement de catégories.

Ci-après se trouvent l'étude des profils lignes et des profils colonnes du premier tableau :

Type	Free to Play	Early Access	Regular Game	
<b>Multijoueur</b>				
Oui	27%	11%	62%	100%
Non	0%	8%	92%	100%
	14%	9%	77%	

FIGURE 3.6 – Profils lignes

En comparant les fréquences marginales, les nombres dévient un peu des chiffres d'ensemble (20% de jeux "Free To Play", 10% de jeux "Early Access", 70% de jeux classiques).

Pour expliquer cela, nous constatons ici que les jeux non multijoueurs sont tous des jeux payants d'accès (Early Access ou Regular Game). Beaucoup de jeux non multijoueurs sont en effet des jeux à scénario, et c'est une forme de jeu qui ne permet économiquement que très peu le format "Free to Play", puisque le jeu en question devrait proposer des bonus internes au jeu et l'intérêt de ceci dans un jeu à scénario est très limité. Ainsi, le mode économique traditionnel est le plus approprié pour les jeux en solo.

La scène multiplayer commence à être enjambée par les jeux "Free To Play", car le format permet facilement aux joueurs actuels d'inviter leurs amis qui ne joueraient pas encore et ainsi de leur permettre de contribuer à la croissance du jeu de manière exponentielle. Selon le type de jeu, il est également facile de proposer aux joueurs des bonus, par exemple de nouveaux niveaux de jeu, des éléments décoratifs en guise de distinction dans le jeu en ligne, etc...

Type	Free to Play	Early Access	Regular Game	
<b>Multijoueur</b>				
Oui	100%	80%	66%	82%
Non	0%	20%	34%	18%
	100%	100%	100%	

FIGURE 3.7 – Profils colonnes

En comparant les fréquences marginales, les nombres dévient un peu des chiffres d'ensemble (26% de jeux solos, 74% de jeux multijoueurs).

La déviation accentue le fait qu'il n'y a pas de jeu solo "Free To Play". Les jeux "Early Access" sont à 80% des jeux multijoueurs : il est en effet judicieux pour l'éditeur de jeux vidéos dans une optique de visibilité de proposer à ses joueurs de tester le jeu avant sa sortie officielle. Nous pouvons ainsi observer ce phénomène, qui en quelques sortes engage les joueurs en tant que "beta-testeurs" et qui rapporte de l'argent et de la popularité à l'éditeur.

Quant au reste des jeux, les jeux multijoueurs font une apparence plus fréquente dans le top 50, mais les jeux solos restent néanmoins des concepts viables économiquement, avec 34% parmi les jeux de configuration économique classique.

Type	Free to Play	Early Access	Regular Game	Sommes
Multijoueur				
Oui	15%	7%	52%	74%
Non	5%	3%	18%	26%
Sommes	20%	10%	70%	100%

Type	Free to Play	Early Access	Regular Game	Sommes
Multijoueur				
Oui	7,4	3,7	25,9	37
Non	2,6	1,3	9,1	13
Sommes	10	5	35	50

FIGURE 3.8 – Profils théoriques (fréquences et effectifs)

En faisant les comparaisons par rapport aux données réelles (cf. Figure 3.5), nous obtenons les résultats suivants :

- Test du khi-deux :  $\chi^2 = 4,855954856$
- Coefficient de contingence :  $CC = 0,297526347$
- V de Cramer :  $V = 0,210107221$

Par les valeurs relativement proches de 0 du coefficient de contingence et du V de Cramer, on peut conclure que le caractère multijoueur et le type économique de jeu sont plutôt indépendants. En effet, autre que le cas particulier du “Free To Play”, les autres paramètres suivent la logique théorique.

### 3.2.2 Quantitatif x qualitatif

Types de jeu	Effectifs	Total Hours Played			Variance totale	Rapport corrélation
		Moyenne	Variance intra	Variance inter		
Free To Play	10	39843117,2	1,88105E+15	8,33375E+14	2,71442E+15	0,307017713
Early Access	5	6117096,8	1,94944E+12	2,29425E+14	2,31374E+14	0,99157449
Regular Game	35	7779993,486	8,65324E+13	2,05885E+14	2,92418E+14	0,704079548
Total	50	14026328,56			2,18E+15	

FIGURE 3.9 – Comparaison heures de jeu et types de jeu

Nous ferons ici l'état du rapport de corrélation  $R = \frac{\text{Var}_{\text{inter}}}{\text{Var}_{\text{totale}}}$ . Plus il est proche de 1, plus les paramètres étudiés sont liés.

Nous obtenons le rapport suivant : les temps de jeux et les types économiques des jeux ne sont que peu liés pour les jeux “Free To Play”, mais nous observons un lien quasiment direct pour les jeux “Early Access” et un lien fort pour les jeux à configuration économique classique.

Le lien pour les jeux “Early Access” par rapport au reste des jeux s'explique temporellement : les jeux “Early Access” sont en effet plus récents que les autres jeux proposés, vu que ce sont des versions beta ou des avant-premières.

Le lien pour les jeux “Regular Game” par rapport au reste s'explique par la philosophie d'acheter un jeu pour une durée de jeu définie. Le type économique proposé propose en effet seulement un ensemble prédéfini et peu extensible de jouabilité et de possibilités pour le consommateur. Il peut éventuellement prolonger la durée de vie de ce jeu en achetant du DLC (*Downloadable Content* ; contenu téléchargeable), mais cela impliquerait de dépenser encore plus que ce qu'il a déjà dépensé.

D'un autre côté, les jeux “Free To Play” sont davantage malléables économiquement : le consommateur n'a rien dépensé encore en commençant, il est donc plus enclin à acheter un peu sur le jeu. Ils sont aussi plus malléables physiquement : augmenter le nombre de fonctionnalités augmente aussi la valeur du jeu, et augmenter le prix d'un jeu à configuration classique n'est pas très judicieux au niveau marketing.

En ce qui concerne les heures de jeu, les jeux “Free To Play” sont soit exploités quotidiennement, perpétuellement (ce qui est souvent le cas pour les jeux stratégiques), soit essayés histoire de “voir ce que ça donne” (car cela ne coûte rien d’essayer, pour un jeu “Free To Play”). Ce qui fait que certains jeux vont avoir plus d’attrait que d’autres, et c’est ce qui donne cette indépendance entre ce type et les heures de jeu.

### 3.2.3 Quantitatif x quantitatif

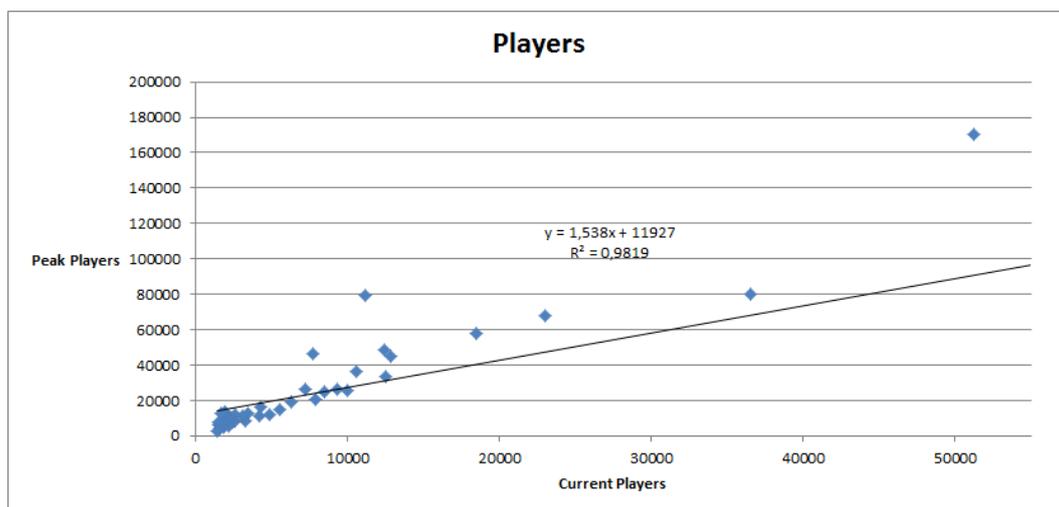


FIGURE 3.10 – Nombre record de joueurs par rapport au nombre de joueurs présents

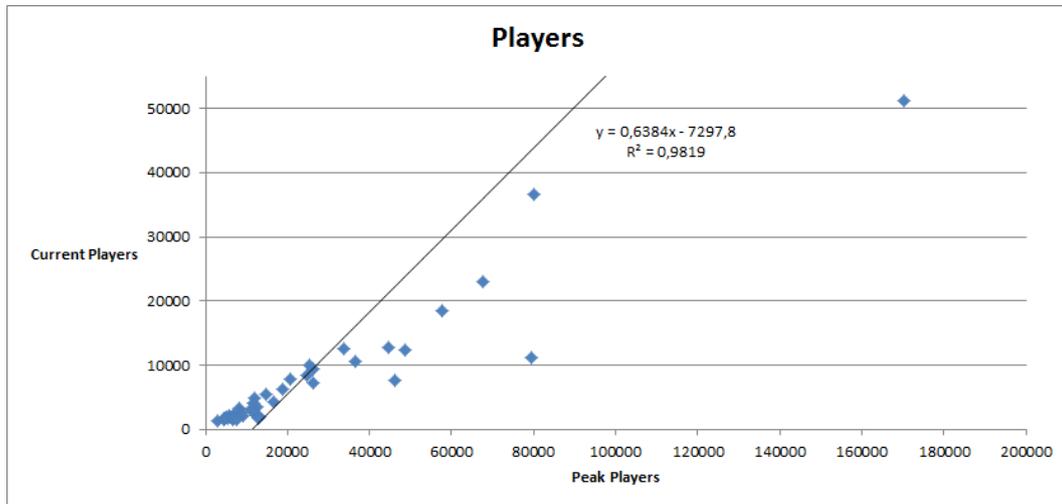


FIGURE 3.11 – Nombre de joueurs présents par rapport au nombre record de joueurs

Nous étudions ici le lien entre le nombre de joueurs présents et le nombre record de joueurs, c'est-à-dire le caractère de la pérennité du succès dont profitent les jeux.

- Moyenne “Current Players” :  $\bar{x} = 17472,76$
- Moyenne “Peak Players” :  $\bar{y} = 38800,76$
- Variance “Current Players” :  $s_x^2 = 5932359308$
- Variance “Peak Players” :  $s_y^2 = 14291939956$
- Covariance :  $c_{xy} = 8941541822$
- Coefficient de corrélation linéaire :  $r_{xy} = \frac{c_{xy}}{s_x s_y} = 0,971075724$

Le coefficient de corrélation linéaire est proche de 1, ce qui veut dire que le rapport entre les deux variables peut se modéliser par une fonction affine. Ce coefficient est positif, donc plus “Peak Players” est grand, plus “Current Players” sera grand.

Ce lien est expliqué par le fait qu'un jeu garde plus ou moins son succès au fil du temps. Par conséquent, son nombre de joueurs présents dépend en effet du nombre record de joueurs qui a pu avoir (le plus souvent le jour de sa sortie).

Nous continuons l'étude des paramètres avec d'autres calculs :

- Erreur globale (variance résiduelle) "Current Players" :  $s_{ex}^2 = 107552977,4$
- Erreur globale (variance résiduelle) "Peak Players" :  $s_{ey}^2 = 259111192,8$
- Variance expliquée "Current Players" :  $s_x^2 = 5824806330$
- Variance expliquée "Peak Players" :  $s_y^2 = 14032828763$
- Coefficient de détermination :  $R^2 = 0,981870117$

Le coefficient de détermination est proche de 1, ce qui dénote que la droite trouvée correspond assez aux données pour ne pas donner trop d'erreurs.

Nous suspectons cependant que les données de "DotA 2" (non présente sur le graphique car trop grande par rapport aux autres) influe sur ce que l'on peut voir sur les figures précédentes, à savoir une déviation visible à l'oeil. Tentons de refaire l'expérience sans cette donnée :

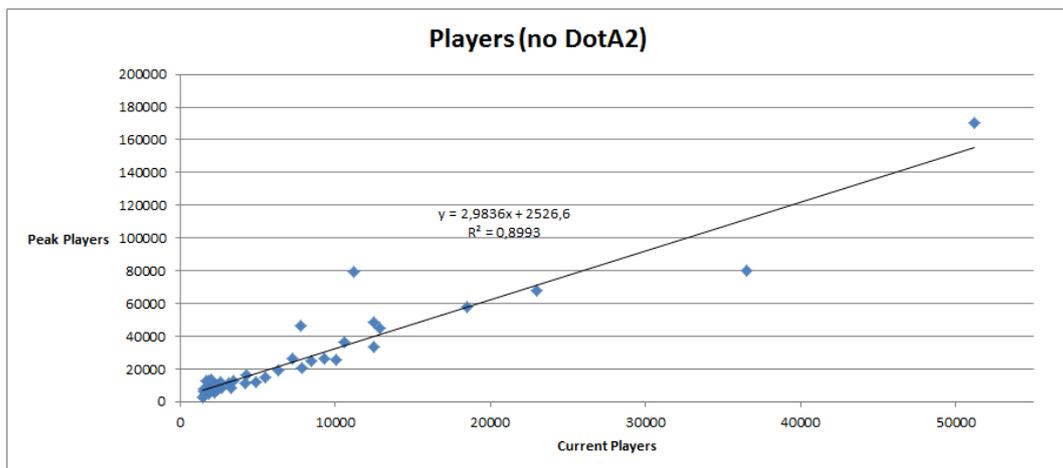


FIGURE 3.12 – Nombre record de joueurs par rapport au nombre de joueurs présents (sans DotA 2)

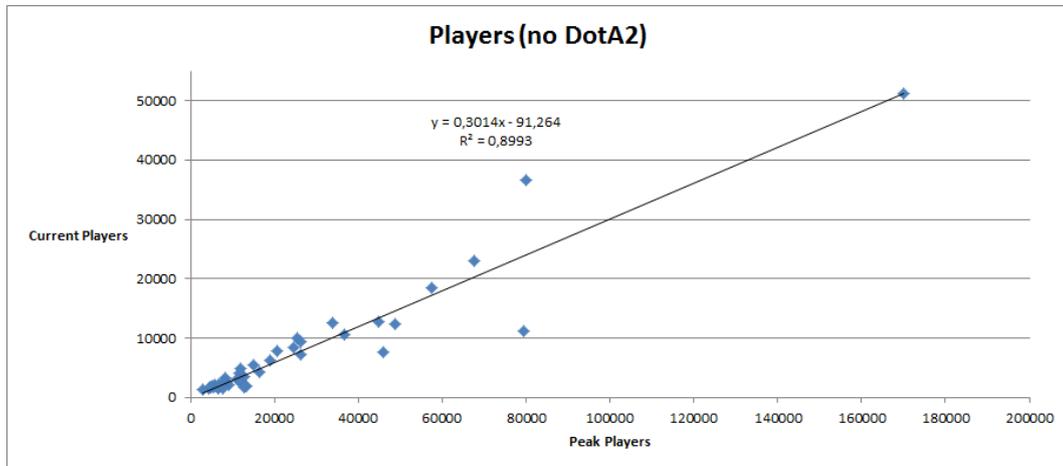


FIGURE 3.13 – Nombre de joueurs présents par rapport au nombre record de joueurs (sans DotA 2)

Nous obtenons ainsi une droite plus cohérente, mais le coefficient de détermination est moins élevé car les erreurs observées se remarquent davantage par les calculs. Cependant,  $R^2 = 0.8993$  reste une valeur correcte pour la modélisation.

Les erreurs visibles par rapport à la droite sont dues aux jeux qui sont soit toujours dans leur phase de succès (nombre de joueurs présents proche du nombre de joueurs record), soit aux jeux qui ont perdu de leur gloire passée (nombre de joueurs présents nettement inférieur au nombre de joueurs record).

# Conclusion

En conclusion, ce travail nous a permis d'observer des éléments de l'industrie du jeu vidéo sur une plateforme en ligne. Les chiffres générés par cette industrie nous ont permis de confirmer ou d'infirmier plusieurs théories que nous nous étions posées à propos des jeux vidéo, notamment des paramètres sociologiques, comme les phénomènes de modes, les conséquences des configurations économiques ou encore comment peut se traduire la durée de vie d'un jeu concrètement.

Il a été intéressant de pouvoir relier des caractères de la vie courante à des chiffres grâce aux outils mis à notre disposition : nous voyons une fois de plus la manifestation des mathématiques, qui a été, est toujours et sera encore omniprésente.