

# METHODOLOGIE : COURBES DE SURVIE ET ESPERANCE DE VIE

En tant qu'experts démographes, vos auteurs et serviteurs n'ont pas résisté à la tentation de calculer des espérances de vie (un démographe ne sachant pas calculer une espérance de vie est un mauvais démographe : un sociologue).

Calculer l'espérance de vie en années des personnages ? « Pfff, trop classique » vous répondront-ils. Ils ont préféré se fixer l'objectif de calculer une espérance de vie en minutes (durée de vie moyenne entre la minute d'apparition du personnage et l'éventuelle minute où celui-ci décède). S'ils y parviennent, ils pourraient ainsi dire qu'un personnage de *Game of Thrones* peut espérer survivre  $x$  minutes. Mieux ! Ils pourraient décliner le calcul selon les caractéristiques des personnages (« un personnage masculin peut espérer survivre  $x$  minutes de plus qu'un personnage féminin »).

## 1. Temps calendaire ou de survie ? Deux approches pour des objectifs différents

Mais comment calculer l'espérance de vie ?

Prenons l'exemple des six personnages suivants : Robb Stark, Daenerys Targaryen, Stannis Barathéon, Brienne de Tarth, Loras Tyrell et Euron Greyjoy (*Fig. 1*).

Bien sûr, ces personnages ne sont pas représentatifs de l'ensemble des personnages nommés : faisant partie des personnages principaux ils sont très loin du personnage normal et peu connu (en plus Brienne est

beaucoup trop cool).

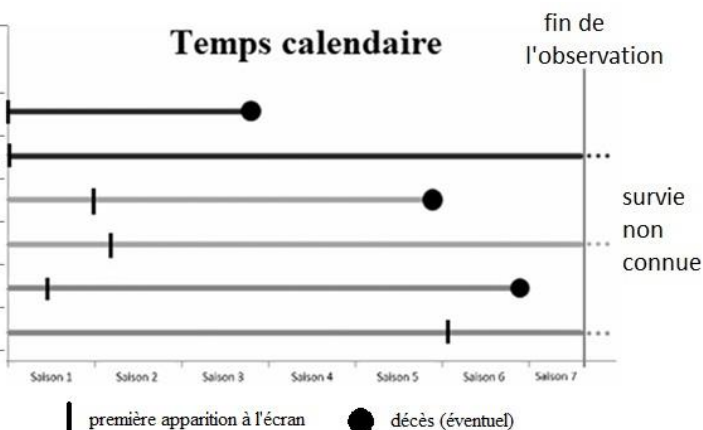
Trois d'entre eux décèdent : Robb, Stannis et Loras – pardon de vous rappeler que nous ne reverrons jamais plus les fesses de Robb. Chronomètre en main, les résultats obtenus par les auteurs pour ces six personnages sont reportés en *Fig. 2*.

Pour calculer la durée de vie moyenne de ces six personnages, certains seraient tentés de faire une moyenne toute simple : en moyenne, ces six personnages vivent 37h et 30min. Pas faux mais cela serait trop facile... et surtout trop biaisé ! La faute d'Euron Greyjoy (entre autres). Celui-ci n'a en effet fait son entrée dans la série qu'à la sixième saison. Par



Fig. 1 : Nos six personnages désignés volontaires pour notre exemple

Première apparition	Dernière apparition	Survie en	Nom	
Episode	Minute	min.		
101	13,5	309 48,9	1482	Robb Stark
101	33,3	707 67,5	3500	Daenerys Targaryen
201	28,5	510 16,2	2011	Stannis Baratheon
203	10,9	707 37,3	2872	Brienne de Tarth
105	7,2	610 17,4	2872	Loras Tyrell
602	40,5	707 22,9	808	Euron Greyjoy



**Lecture :** Stannis Barathéon apparaît à la 28ème minute du 1er épisode de la saison 2. Il meurt à la 16ème minute du dernier épisode de la cinquième saison. Entre ces deux moments, il s'est écoulé (hors générique) 2 011 minutes. Stannis a donc vécu l'équivalent de 33 heures et 31 minutes : admirable !

Fig. 2 : Temps de survie calendaire

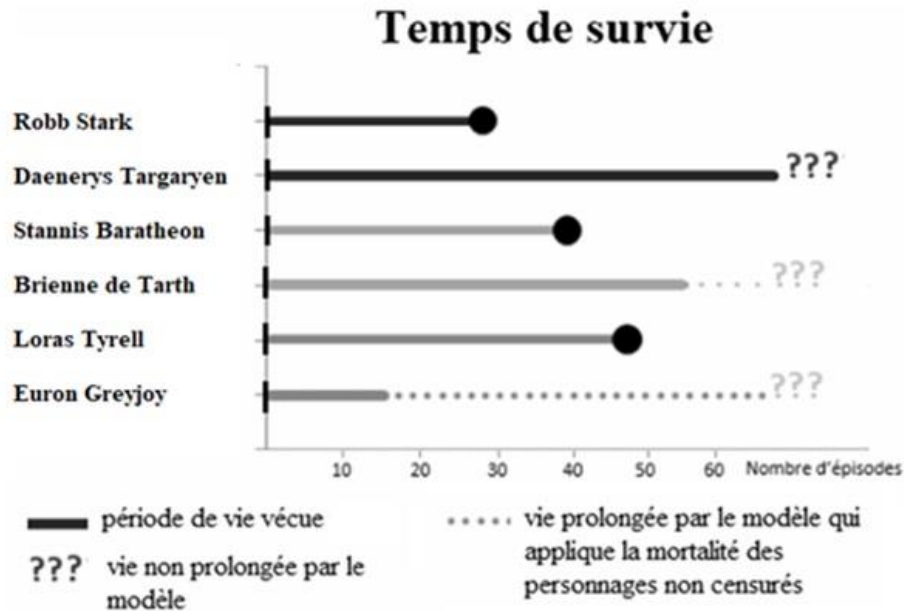


Fig. 3 : Temps de survie avec horloges synchronisées

conséquent, il n'a eu à survivre *qu'*à deux saisons. Est-il donc juste de dire que son espérance de vie était faible ? Non. Et d'ailleurs, à cause de lui, la moyenne est biaisée car tirée vers le bas.

Vous l'aurez compris (et si ce n'est pas le cas on vous l'affirme) le calcul de la durée moyenne de survie est compliqué par le fait que l'on est incapable d'observer les personnages sur une même durée (ça aurait été vachement cool de suivre Lyanna Mormont pendant 67 épisodes). Alors comment faire ?

Faisons comme si tous les personnages étaient apparus en même temps dans la série. On dit que l'on « synchronise les horloges » (Fig. 3).

Accrochez-vous, vous avez fait le plus dur... Ou presque. Reprenons l'exemple d'Euron Greyjoy. Nous ne savons pas ce qu'il advient de lui après 17 épisodes. C'est là que le modèle de durée non paramétrique intervient (de son petit nom... Kaplan-Meier).

Une fois les horloges synchronisées, le modèle permet d'appliquer à Euron les risques de mourir de Loras, Robb, Daenerys, et de Brienne que l'on suit durant plus de 17 épisodes. Le modèle est donc en mesure d'estimer les risques qu'a un personnage de mourir compte tenu de ce que nous avons observé chez d'autres personnages. Les sept

premières saisons de *Game of Thrones* durent 65 heures. Passée cette durée, il nous est impossible d'appliquer ce modèle puisqu'aucun personnage n'est observé plus longtemps.

Finalement, quel genre de résultats obtient-on ? Celui-ci, par exemple : l'espérance de vie des personnages féminins est de 29 heures contre 22 heures et 30 minutes pour les personnages masculins. Ces quelques chiffres ne vous suffisent pas ? N'hésitez pas à consulter le dossier de « suivi des biographies » qui en regorge !

## 2. Courbes de survie

Vous êtes-encore là vous ? Coriaces... Nous allons tenter de vous assommer définitivement.

Donc, l'espérance de vie ça vous parle maintenant. Par contre, les courbes de survie que vous avez croisées sur notre site ou dans nos articles (dossier de « suivi des biographies », dossier « cohorte », l'article « Une cohorte plus à l'abri ? », le dictionnaire des variables) ça, vous ne connaissez pas encore. Normal, on entre encore un peu plus dans la tambouille du démographe ! Rien de tel qu'un exemple.

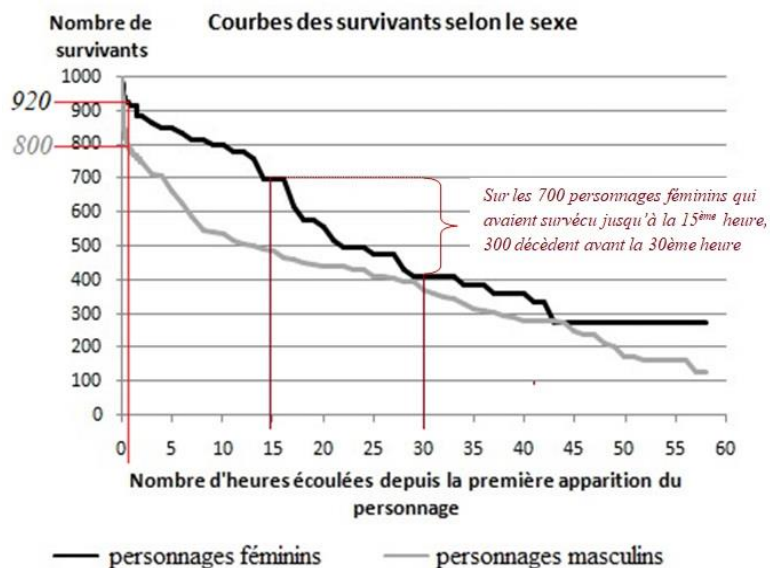


Fig. 4 : Courbes de survie des personnages masculins et féminins

Tout à l'heure, nous avons calculé un risque de mourir pour chaque heure de survie (mais siii, grâce au modèle de Kaplan-Meier).

La courbe de survie découle de ce premier calcul. Pour comprendre la courbe de survie, il faut faire comme si nous suivions dès la première minute (synchronisation des horloges oblige) 1000 personnages masculins et 1000 personnages féminins. Et après ? Bah c'est simple ! On fait ce que l'on fait de mieux ! On regarde les personnages mourir les uns après les autres en mangeant du popcorn.

Concrètement, comment se passe le calcul ? Si je vous dis qu'un personnage féminin a 30% de risque de mourir au cours des 15 premières heures de survie, combien de morts cela fait-il ? 30% de risque de mourir fois 1000 personnages = 300 personnages féminins en moins sur les 15 premières heures ! Après 15h, il ne reste donc plus que 700 survivantes théoriques (Fig. 4). Au passage, nous parlons de survivants théoriques car nous partons de 1000 personnages fictifs. Mais ça, c'est de la poésie !

Alors ça y est, vous l'avez ? Un dernier calcul pour en être certain. Au cours des 15 heures de survie qui suivent, un personnage féminin a 43% de risque de mourir. 43% fois 700 personnages encore en vie = 300 morts

supplémentaires entre la quinzième et la trentième heure de survie (Fig. 4).

Voilà, vous avez compris ! Maintenant il ne vous reste plus qu'à vous amuser en comparant les courbes de survie des personnages féminins et des personnages masculins. Sur la première heure de survie, la différence de mortalité est impressionnante : 8% des personnages féminins décèdent (il reste environ 920 survivantes) contre plus de 20% des personnages masculins (il reste 800 survivants).

Mais à la trentième heure de survie, il reste quasiment autant de personnages masculins que de personnages féminins. Les courbes de survie nous renseignent donc bien sur l'évolution des risques de mortalité dans le temps.

Afin de déterminer si deux courbes de survie sont différentes entre elles, nous nous appuyons sur la statistique du *LogRank* et de celle de *Wilcoxon*. Que de noms aussi barbares que les *Dothrakis* ! Si les auteurs se gardent bien de vous expliquer ces tests, c'est parce qu'ils ne les comprennent pas dans le détail. Ils se contentent d'étudier le résultat calculé par l'ordinateur. Celui-ci sort un pourcentage qui est le risque à prendre pour rejeter l'égalité entre les courbes de survie. Par exemple, si le risque est de 4%, on peut affirmer que les courbes de survie sont différentes : juré craché, nous en sommes certains... A 96% ! Pour différencier ou non les courbes, le test de *Wilcoxon* accorde plus d'importance aux décès précoces. Il nous sera utile puisque de nombreux personnages décèdent rapidement dans *Game of Thrones*.

Ce modèle non paramétrique a toutefois une limite majeure. Il se concentre sur l'impact des modalités d'une variable (personnage masculin/féminin) sans prendre en compte l'influence potentielle d'autres variables (les personnages masculins sont en proportion plus souvent des combattants tandis que les personnages féminins sont plus souvent des

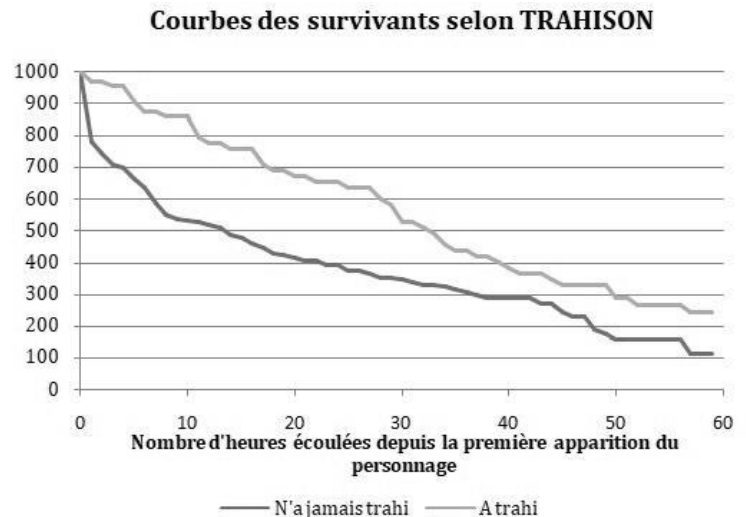
prostituées : on parle d'effets de composition). Pour une analyse plus fine, l'emploi de régressions logistiques est alors nécessaire.

### 3. Les courbes de survie prédéterminées

Nous pouvons juger que certaines variables prédéterminent le futur des personnages : un personnage qui a trahi a dû vivre en moyenne plus longtemps pour le faire (de la même manière pour un personnage qui a tué par exemple). Le personnage ne vit pas longtemps parce qu'il a tué ou trahi mais il a tué ou trahi parce qu'il en a eu le temps et parce qu'il a vécu longtemps (Fig. 5).

Ainsi, nous ne pouvons analyser correctement de telles courbes de survivants car celles-ci n'ont pas de sens. En effet, la démarche est aussi absurde que celle d'un démographe qui ferait deux courbes de survie : une pour les personnes qui ont vécu en maison de retraite, une pour celles qui n'y sont jamais allées. Il en résultera automatiquement que les personnes qui sont allées en maison de retraite dans leur vie auront une espérance de vie plus longue mais quel sens donner à ce résultat ?

Par contre la majorité des variables prédéterminées pourra être exploitée dans les régressions logistiques. En effet, nous pouvons grâce aux régressions logistiques raisonner à épisode de vie égal. A épisode de vie égal, avoir trahi ou tué impacte-t-il la survie au cours d'un épisode ?



**Lecture :** Ceux qui ont trahi au moins une fois vivent en moyenne plus longtemps (car ils ont eu le temps de trahir) ; analyse impossible car il s'agit d'une courbe de survie prédéterminée. Seuils inférieurs à 0,01% (d'après statistiques du Log-Rang et de Wilcoxon)

Fig. 5 : Courbes de survie selon la variable trahison

### 4. Résumé des méthodologies

Modèle Logit	Modèle Kaplan-Meier	
Régression logistique	Espérance de vie	Courbes de survie
« Si personnages féminins et personnages masculins avaient les mêmes caractéristiques, auraient-ils un risque de décéder différent ? »	« Combien de temps un personnage survit-il en moyenne à partir de son apparition à l'écran <sup>1</sup> ? »	« Comment varient les risques de décès des personnages au cours du temps ? »
Calcule des <i>Odds Ratios</i> qui permettent d'établir les liens « purs » entre les modalités d'une variable et le risque de mourir <i>toutes choses égales par ailleurs</i> . Calcule une probabilité de décès pour <b>chaque personnage</b> .	Calcule une durée de survie moyenne pour des <b>groupes d'individus</b> . Limite : ne tient pas compte des effets de composition	Trace une courbe de survie pour des <b>groupes d'individus</b> qui permet de suivre l'évolution des risques de décès au cours du temps.

#### Bibliographie

Alberti, C. et all (2005) « Analyse de survie : le test du logrank ». [En ligne] : [http://www.biomedicale.parisdescartes.fr/enseignement/toxico/M2THERV\\_2013\\_2014/documents/C1/Rou dot%20logrank.pdf](http://www.biomedicale.parisdescartes.fr/enseignement/toxico/M2THERV_2013_2014/documents/C1/Rou dot%20logrank.pdf) [Consulté le 13 avril 2018]

Laporte, S. (2005) « Comment lire une courbe de survie ? ». In : *mt*. Vol. 11, n° 6, pp419-423.