

ETUDE STATISTIQUE DU FICHIER TITANIC



Etudiant : Ismail IBRAHIM

N° : 42221614

## INTRODUCTION :

Ce projet d'étude statistique a pour objectif d'analyser les données des passagers du Titanic pour identifier les facteurs qui ont influencé les chances de survie lors de cette tragédie. En examinant des variables telles que la classe de voyage, le sexe, l'âge, et le point d'embarquement, nous chercherons à dégager des tendances significatives.

Grâce à des techniques d'analyse descriptive et de visualisation graphique, cette étude mettra en lumière les différents critères qui ont pu jouer un rôle crucial dans la survie des passagers.

### Analyse descriptive

Pclass			Sex			Embarked		
Pclass	Fréquence	Pourcentage	Sex	Fréquence	Pourcentage	Embarked	Fréquence	Pourcentage
1	214	24.21	female	312	35.29	C	167	18.93
2	184	20.81	male	572	64.71	Q	76	8.62
3	486	54.98				S	639	72.45
						Fréquence manquante = 2		

Après avoir mené des analyses descriptives, nous pourrions en déduire qu' il y a 214 passagers dans la classe 1 soit 24.21 %, 184 passagers dans la classe 2 soit 20.81 % et 486 passagers dans la classe 3 soit environ plus de la moitié de tous les passagers (54.98 %).

Et parmi ces passagers, il y a 312 femmes et 572 hommes soient respectivement 35.29 % et 64.71 %.

Les données montrent que les passagers du Titanic ont embarqué depuis trois ports distincts : Cherbourg, Queenstown et Southampton. Le nombre de passagers embarqués dans chacun de ces ports varie considérablement : **Cherbourg** : 167 passagers soient environ 18.93 % ;

**Queenstown** :76 passagers soient 8.62 % et en fin **Southampton** : qui comptabilise la majorité de passagers soient 639 passagers environ 72.45 % de tous les passagers car Southampton était un port de départ majeur pour les traversées transatlantiques. De plus, étant le port de départ initial du Titanic, il est logique qu'une majorité des passagers aient embarqué ici.

**Cherbourg**, était une escale importante pour ceux qui souhaitaient rejoindre le Titanic sans devoir voyager jusqu'à Southampton. Bien que le nombre de passagers soit inférieur à celui de Southampton, il reste significatif.

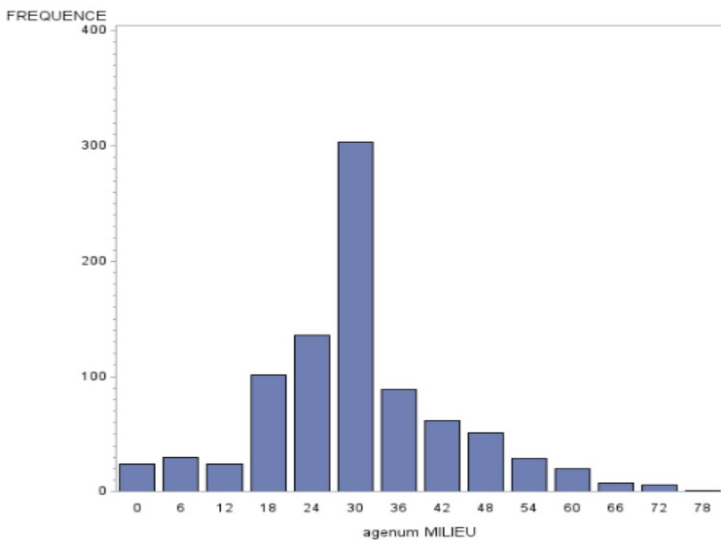
**Queenstown** était l'escale finale avant que le Titanic ne traverse l'Atlantique ce qui explique son infériorité numérique par rapport aux autres ports. En résumé, cette répartition souligne l'importance logistique et stratégique des ports d'embarquement dans l'organisation des traversées transatlantiques à l'époque du Titanic.

Survived		
Survived	Fréquence	Pourcentage
0	543	61.43
1	341	38.57

Les données révèlent un triste constat : sur les passagers du Titanic, 543 n'ont pas survécu à la tragédie, tandis que 341 ont réussi à survivre. Cela signifie que seulement environ 38.57% des passagers ont survécu, tandis qu'une majorité de 61.43% n'ont pas survécu.



## INTERPRETATION GRAPHIQUE



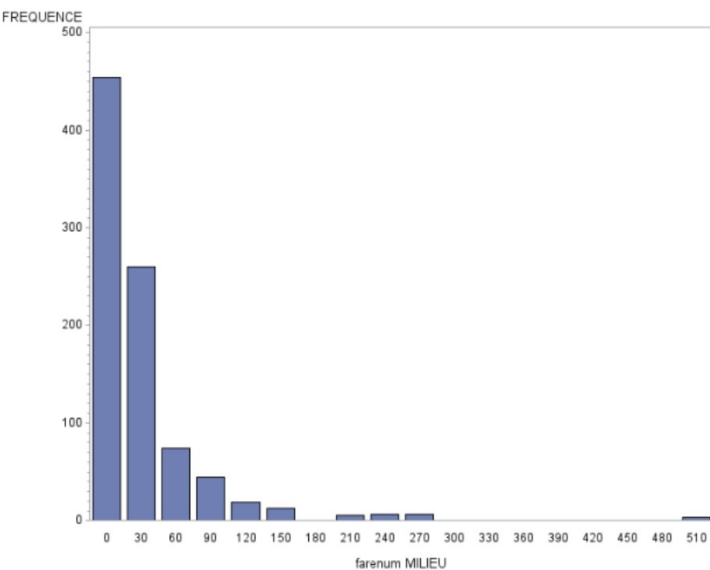
L'âge des passagers varie de 0 à 78 ans. La distribution des âges semble suivre une forme approximativement normale, avec une seule grande concentration autour de la valeur centrale.

L'âge le plus fréquent (mode) semble se situer autour de 30 ans, avec une fréquence légèrement supérieure à 300 passagers.

Cela indique que la majorité des passagers étaient autour de cet âge.

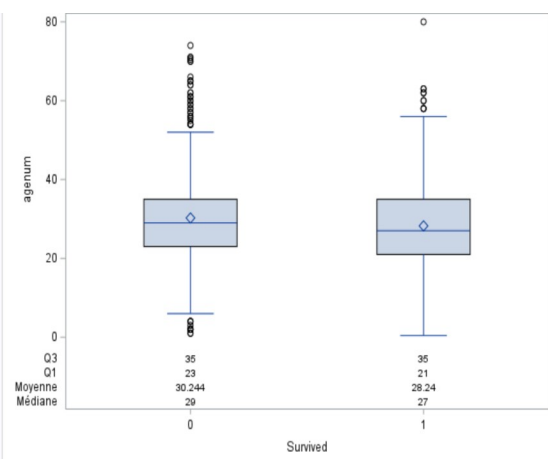
La fréquence des âges diminue progressivement de part et d'autre de la concentration centrale, formant une distribution symétrique. Cette forme suggère qu'il y avait moins de très jeunes et de très vieux passagers par rapport à ceux d'âge moyen. Les

jeunes passagers (âgés de 0 à 20 ans) représentent un nombre significatif, mais moindre que le groupe central. Les passagers plus âgés (plus de 50 ans) sont également présents, mais en moins grand nombre. L'histogramme de la variable Âge est essentiel pour comprendre la démographie des passagers du Titanic. Il montre clairement que la majorité des passagers étaient des adultes d'âge moyen, avec une représentation moindre des enfants et des personnes âgées



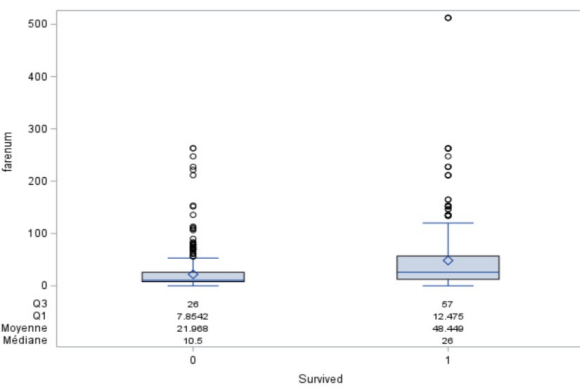
L'histogramme de la variable Fare (tarif du ticket) affiche la distribution des tarifs payés par les passagers du Titanic. La majorité des tarifs des tickets sont concentrés dans la tranche de prix la plus basse. Le pic de fréquence le plus élevé se trouve dans le premier intervalle (0-30), avec une fréquence d'environ 450 passagers. L'intervalle suivant (30-60) présente une fréquence d'environ 260 passagers. Les intervalles suivants montrent une diminution rapide de la fréquence, avec l'intervalle 60-90 ayant une fréquence d'environ 80 passagers. Les fréquences continuent de diminuer de manière significative pour les tranches de tarifs plus élevés, avec très peu de passagers ayant payé

plus de 150. Cependant, il y a un petit pic de fréquence dans la tranche de tarifs la plus élevée (480-510), indiquant quelques rares occurrences de tickets très chers. Cette distribution montre que la plupart des passagers voyageaient avec des billets à bas prix, ce qui suggère une prédominance de passagers de troisième classe ou ceux ayant des moyens financiers modestes. Les rares occurrences de tarifs élevés correspondent probablement aux passagers de première classe, soulignant les disparités socio-économiques présentes à bord du Titanic. L'histogramme de la variable fare met en évidence la forte concentration de passagers ayant payé des tarifs relativement bas pour leurs tickets. Cette visualisation permet de voir clairement que bien que quelques passagers aient payé des sommes considérables, la grande majorité voyageait avec des billets bien plus abordables.



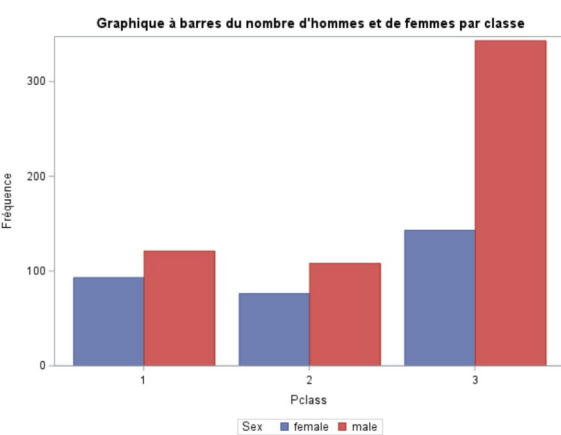
La médiane (âge central) des non-survivants est légèrement plus élevée (29 ans) que celle des survivants (27 ans). De même, la moyenne d'âge des non-survivants (30,24 ans) est légèrement supérieure à celle des survivants (28,24 ans). Cela suggère que les survivants étaient en moyenne légèrement plus jeunes que les non-survivants. Les premier et troisième quartiles sont assez similaires pour les deux groupes, ce qui signifie que la majorité des âges des deux groupes se chevauchent, avec des âges compris entre environ 21 et 35 ans pour les survivants et 23 à 35 ans pour les non-survivants. Les non-survivants sont plus présents dans les tranches d'âge les plus élevées, suggérant que les passagers plus âgés avaient moins de chances de survivre. Cette boîte à moustaches

indique que bien que les âges des passagers soient globalement similaires entre les survivants et les non-survivants, il existe une légère tendance où les jeunes passagers ont eu de meilleures chances de survie. La présence des valeurs extrêmes dans les âges élevés parmi les non-survivants renforce l'idée que l'âge avancé pouvait être un facteur de risque plus élevé lors de cette tragédie.



La boîte à moustaches de la variable fare (tarif du ticket) en relation avec la variable survived (survie) présente une comparaison entre les passagers qui ont survécu et ceux qui n'ont pas survécu. Plusieurs valeurs extrêmes avec des tarifs atteignant environ 300. Les passagers qui ont survécu ont payé en moyenne des tarifs plus élevés (médiane de 26 et moyenne de 48,448) par rapport à ceux qui n'ont pas survécu (médiane de 10,5 et moyenne de 21,968). Cela suggère que les passagers ayant payé des tarifs plus élevés avaient de meilleures chances de survie, probablement parce qu'ils étaient en première ou deuxième

classe. Les valeurs extrêmes indiquent qu'il y avait des passagers avec des tarifs très élevés dans les deux groupes. Cependant, il y a une plus grande concentration de tarifs élevés parmi les survivants, ce qui renforce l'hypothèse que la classe de voyage et la capacité à payer des tarifs plus élevés étaient associées à de meilleures chances de survie. Cette boîte à moustaches met en évidence une corrélation entre le tarif du ticket et les chances de survie lors du naufrage du Titanic. Les passagers ayant payé des tarifs plus élevés, souvent associés aux classes supérieures, avaient une probabilité accrue de survie.



Le graphique à barres montre la distribution des hommes et des femmes dans les différentes classes de voyage (1ère, 2ème et 3ème classe) à bord du Titanic.

► Il y a plus d'hommes que de femmes dans la première classe, Bien que moins nombreuses que les hommes, une proportion significative de femmes voyageait en première classe.

► La deuxième classe compte également plus d'hommes que de femmes, mais la différence est légèrement moins marquée que dans la première classe. Le nombre de femmes est encore une fois inférieur à celui des hommes, mais représentatif de la classe

moyenne.

► La troisième classe présente une nette prédominance d'hommes, avec une fréquence dépassant largement les 300. Les femmes sont significativement moins nombreuses en troisième classe comparativement aux hommes. La distribution des sexes par classe peut avoir un impact direct sur les chances de survie, car les politiques de sauvetage donnaient la priorité aux femmes et aux enfants. La présence de plus de femmes en première classe pourrait avoir contribué à des taux de survie plus élevés pour cette classe.

## Analyse statistique et conclusion

Nombre de passagers adultes et d'enfants

La procédure FREQ

category	Fréquence	Pourcentage
Adulte	772	87.33
Enfant	112	12.67

Ce tableau affiche la répartition des passagers adultes et enfants du Titanic. Il y a 772 passagers adultes (87.33% du total) et 112 enfants passagers (12.67% du total). La majorité des passagers étaient des adultes à bord du Titanic.

Nombre d'enfants qui ont survécu à la tragédie

Obs.	category	Survived	COUNT
1	Enfant	1	61

Ce tableau montre que parmi les enfants présents sur le Titanic, 61 ont survécu à la tragédie.

Nombre d'adultes qui ont survécu à la tragédie

Obs.	category	Survived	COUNT
1	Adulte	1	280

Ce tableau montre que parmi les passagers adultes présents sur le Titanic, 280 ont survécu à la tragédie. Cette donnée met en évidence un nombre significatif d'adultes survivants, ce qui peut être attribué à plusieurs facteurs comme : Priorité de Sauvetage ; Accès aux Canots de Sauvetage ou Capacité de Survie (avantages physiques).

Obs.	Pclass	category	Embarked	COUNT
1	1	Adulte		2
2	1	Adulte	C	55
3	1	Adulte	Q	1
4	1	Adulte	S	66
5	1	Enfant	C	3
6	1	Enfant	S	8
7	2	Adulte	C	5
8	2	Adulte	Q	2
9	2	Adulte	S	59
10	2	Enfant	C	4
11	2	Enfant	S	17
12	3	Adulte	C	14
13	3	Adulte	Q	24
14	3	Adulte	S	52
15	3	Enfant	C	11
16	3	Enfant	Q	3
17	3	Enfant	S	15

Ce tableau montre le nombre d'enfants et d'adultes qui ont survécu à la tragédie du Titanic, répartis par classe et par point d'embarquement. **En 1ere CLASSE :** Les adultes embarqués à Cherbourg et Southampton avaient des taux de survie relativement élevés, probablement en raison de l'accès plus facile aux canots de sauvetage. Les enfants en première classe avaient également de bonnes chances de survie, avec 11 survivants.

**En 2emeCLASSE:** La majorité des survivants adultes de la deuxième classe ont embarqué à Southampton, montrant une meilleure préparation et peut-être un accès plus favorable aux ressources de sauvetage. Les enfants de cette classe avaient également un taux de survie élevé, surtout ceux embarqués à Southampton. **En 3emeCLASSE:** Les passagers adultes de troisième classe embarqués à Southampton ont survécu en plus grand nombre, bien que la majorité ait embarqué à Queenstown et

Southampton. Les enfants en troisième classe ont également eu une chance de survie notable, particulièrement ceux embarqués à Southampton. **Les passagers ayant embarqué à Southampton, en particulier ceux en première et deuxième classe, avaient une meilleure probabilité de survie. Les enfants de toutes les classes, bien que moins nombreux, avaient des chances de survie significatives, surtout ceux en première et deuxième classe, démontrant l'impact des priorités de sauvetage.**

## CONCLUSION

La majorité des passagers voyageaient en troisième classe, suivis de la deuxième et de la première classe et la plupart des passagers étaient des adultes, avec une minorité d'enfants. Les hommes étaient plus nombreux que les femmes dans chaque classe, surtout en troisième classe. Les passagers de première classe avaient les meilleures chances de survie, en grande partie grâce à leur accès privilégié. Les passagers de deuxième classe avaient des chances de survie raisonnables, mais inférieures à celles de la première classe. Les passagers de troisième classe avaient les plus faibles chances de survie en raison de leur accès limité aux équipements de sauvetage et de leur position géographique dans le navire. Les passagers embarqués à **Southampton** avaient les meilleures chances de survie, suivis de ceux embarqués à **Cherbourg** et **Queenstown**.

## ANNEXE

```
libname projet "C:\Users\ibrah\Documents"; run;
/*etape importation des fichiers*/
PROC IMPORT
    DATAFILE= "C:\Users\ibrah\Documents\Projet SAS\train_url1.xlsx"
    OUT= Projet.train_url1 DBMS=xlsx
    REPLACE; SHEET="train_url1"; GETNAMES=YES; RUN;
PROC IMPORT
    DATAFILE= "C:\Users\ibrah\Documents\Projet SAS\train_url2.xlsx"
    OUT= Projet.train_url2
    DBMS=xlsx
    REPLACE;
    SHEET="Feuil1";
    GETNAMES=YES; RUN;
PROC IMPORT
    DATAFILE= "C:\Users\ibrah\Documents\Projet SAS\train_url3.xlsx"
    OUT= Projet.train_url3
    DBMS=xlsx
    REPLACE;
    SHEET="Feuil2";
    GETNAMES=YES; RUN;
proc sort data=Projet.train_url1;
by PassengerId; run;
proc sort data=Projet.train_url2;
by PassengerId; run;
/*1.Concatenation Horizontale*/
DATA Projet.train_url1_2;
    MERGE Projet.train_url1 Projet.train_url2 ;
    BY PassengerId; RUN;
proc print data=Projet.train_url1_2;
run;
/*Concatenation verticale*/
DATA projet.train_url;
    SET Projet.train_url1_2 projet.train_url3; RUN;
proc print data=projet.train_url; run;
/*Changement de Age et Fare en numerique*/
data projet.train_url;
set projet.train_url;
agenum=input(age, best4.);
farenum=input(fare, best8.);
drop age fare; run;
proc contents data=projet.train_url; run;
proc print data=projet.train_url; run;
/*Etape 2 : Analyse descriptive du fichier Train_url */
/*1. Description de toutes les variables présentes dans le fichier train_url*/
PROC CONTENTS DATA=projet.train_url;
Run;
/*2.les statistiques descriptives pour toutes les variables*/
proc means data=projet.train_url n nmiss mean median std min max;
var Survived Pclass Agenum Farenum;
label mean = "Moyenne" median = "Médiane" std = "Écart type" min = "Minimum"
max = "Maximum"; run;
proc freq data=projet.train_url;
tables Sex Embarked / nocum; run;
/*3. Combien y a-t-il de passagers par classe ? */
proc freq data=projet.train_url;
tables pclass / nocum; run;
/*4. Combien y a-t-il de passagers hommes et femmes ? */
proc freq data=projet.train_url;
tables sex / nocum; run;
/*5. Combien y a-t-il de passagers par point d'embarquement ? */
proc freq data=projet.train_url;
```

```

    tables embarked / nocum;run;
/*6. Déterminer le nombre de survivants et le nombre de non survivant ?*/
proc freq data=projet.train_url;
    tables survived / nocum;run;
/*Etape 3 : graphiques et agrégations */
/*a. Faire un histogramme de la variable Age. Interprétez le graphique*/
PROC GCHART DATA=projet.train_url;
    VBAR agenum;RUN;
/*b. Faire une boîte à moustaches de Age par rapport à la survival*/
proc sgplot data=projet.train_url;vbox agenum / category=Survived
displaystats=(median mean q1 q3);run;
/*c. Faire un histogramme de la variable Fare. Interprétez le graphique */
PROC GCHART DATA=projet.train_url;VBAR farenum;RUN;
/*d.une boîte moustaches de la variable Fare par rapport à survival*/
proc sgplot data=projet.train_url;vbox farenum / category=Survived
displaystats=(median mean q1 q3);run;
/*a. Écrivez un script pour calculer la moyenne de la variable Âge par sexe. */
proc means data=projet.train_url mean;class sex;var agenum;run;
/*b.la moyenne par classe des variables Fare et Age */
proc means data=projet.train_url mean;class pclass;var Farenum Agenum;run;
/*c.le nombre d'hommes et de femmes par classe*/
proc freq data=projet.train_url;tables pclass*sex / norow nocol nopercnt;run;
proc sgplot data=projet.train_url;vbar pclass / group=sex
groupdisplay=cluster;run;
/*d. la moyenne de la variable Age par classe et par sexe */
proc means data=projet.train_url mean;class pclass sex;var Agenum;run;
/*e. Déterminer et remplacer toutes les valeurs manquantes. */
proc means data=projet.train_url n nmiss mean median;var agenum;class sex;run;
data projet.train_url;set projet.train_url;if sex='male' and agenum=. then
agenum = 29;if sex='female' and agenum=. then agenum = 27;run;
proc means data=projet.train_url nmiss;var agenum;run;
proc print data=projet.train_url;run;
/*Etape 4 : Analyse statistique et conclusion : */
/*A. Créer une variable pour différencier un adulte d'un enfant */
data nouvelle_variable;set projet.train_url;if agenum < 18 then category =
'Enfant';else category = 'Adulte';run;
proc print data=nouvelle_variable;run;
/*B. le nombre de passagers adultes et le nombre d'enfants passagers */
proc freq data=nouvelle_variable;tables category / nocum;run;
/*C. Combien d'enfants ont survécu à la tragédie ? */
proc freq data=nouvelle_variable;tables category*survived / noprint
out=counts;run;
data enfants_survivants;set counts;if category = 'Enfant' and survived = 1;run;
proc print data=enfants_survivants;var category survived count;run;
/*D. Combien d'adultes ont survécu à la tragédie ?*/
proc freq data=nouvelle_variable;tables category*survived / noprint
out=counts;run;
data adultes_survivants; set counts;if category = 'Adulte' and survived = 1;run;
proc print data=adultes_survivants;var category survived count;run;
/*E. Quel est le nombre de survivants par classe ? */
proc freq data=nouvelle_variable;tables pclass*survived / noprint
out=survivants_par_classe;run;
proc print data=survivants_par_classe;var pclass survived count;where survived =
1;run;
/*F. Combien d'enfants et d'adultes ont survécu par classe ?*/
proc freq data=nouvelle_variable;tables pclass*category*survived / noprint
out=survivants;run;
data survivants;set survivants;if survived = 1;run;
proc print data=survivants;var pclass category count;run;
/*G.enfants et d'adultes qui ont survécu par classe et par point d'embarcation*/
proc freq data=nouvelle_variable;tables pclass*category*Embarked*survived /
noprint out=survivant_embarquement;run;
data survivants;set survivant_embarquement;if survived = 1;run;
proc print data=survivants;var pclass category Embarked count;run;

```