



Statistique exploratoire

Thierry Dhorne

1^{er} mars 2016



Introduction

- ❖ Présentation
- ❖ Programme
- ❖ Intérêt
- ❖ The german tank problem
- ❖ Problème du char d'assaut allemand
- ❖ Contexte

Statistique exploratoire

Bases (et vocabulaire) de la statistique (exploratoire)

Introduction



Présentation

Introduction

❖ Présentation

❖ Programme

❖ Intérêt

❖ The german tank problem

❖ Problème du char d'assaut allemand

❖ Contexte

Statistique exploratoire

Bases (et vocabulaire) de la statistique (exploratoire)

- Thierry Dhorne
- Professeur de Statistique
- thierry.dhorne@univ-ubs.fr
- www.dhorne.education



Programme

Introduction

❖ Présentation

❖ Programme

❖ Intérêt

❖ The german tank problem

❖ Problème du char d'assaut allemand

❖ Contexte

Statistique exploratoire

Bases (et vocabulaire) de la statistique (exploratoire)

- Statistique (exploratoire)
 - Probabilités (à venir)
 - Statistique (inférentielle) : l'année prochaine



Intérêt

Introduction

- ❖ Présentation
- ❖ Programme

❖ Intérêt

- ❖ The german tank problem
- ❖ Problème du char d'assaut allemand
- ❖ Contexte

Statistique exploratoire

Bases (et vocabulaire) de la statistique (exploratoire)

- culture et compétence scientifique
- société numérique - cyber (data science)
- statistique militaire : le problème du panzer !



The german tank problem

Introduction

- ❖ Présentation
- ❖ Programme
- ❖ Intérêt
- ❖ The german tank problem
- ❖ Problème du char d'assaut allemand
- ❖ Contexte

Statistique exploratoire

Bases (et vocabulaire) de la statistique (exploratoire)

The screenshot shows the Wikipedia article for 'German tank problem'. At the top right, there are links for 'Create account' and 'Log in'. Below these are navigation tabs for 'Article' and 'Talk', and a search bar. The main heading is 'German tank problem', followed by the text 'From Wikipedia, the free encyclopedia'. The article text begins with 'In the statistical theory of estimation, the problem of estimating the maximum of a discrete uniform distribution from sampling without replacement is known in English as the **German tank problem**, due to its application in World War II to the estimation of the number of German tanks.' It then discusses the difference between frequentist and Bayesian inference. A 'Contents' table of contents is visible, listing sections from 'Example' to 'References'. To the right of the text is an image of German tanks in a field, with a caption stating: 'During World War II, production of German tanks such as the Panther was accurately estimated by Allied intelligence using statistical methods.' On the left side of the page, there is a sidebar with the Wikipedia logo and various navigation links such as 'Main page', 'Contents', 'Featured content', 'Random article', 'Donate to Wikipedia', 'Wikimedia Shop', 'Interaction', 'Help', 'About Wikipedia', 'Community portal', 'Recent changes', 'Contact page', 'Tools', 'Print/export', 'Languages', and 'Edit links'.



Problème du char d'assaut allemand

Introduction

- ❖ Présentation
- ❖ Programme
- ❖ Intérêt
- ❖ The german tank problem

❖ Problème du char d'assaut allemand

❖ Contexte

Statistique exploratoire

Bases (et vocabulaire) de la statistique (exploratoire)

The screenshot shows the Wikipedia article page for 'Problème du char d'assaut allemand'. The page includes a navigation bar at the top with user options like 'Discussion', 'Brouillon', and 'Préférences'. The article title is prominently displayed. Below the title, there is a summary paragraph explaining that the problem refers to estimating the maximum value of a uniform distribution from a sample without replacement. A mathematical formula is provided:
$$N \approx m - 1 + \frac{m}{k}$$
 The article also features a table of contents with sections for 'Histoire', 'Autres applications', 'Notes et références', 'Bibliographie', and 'Voir aussi'. Two images are included: one showing a line of Panzerkampfwagen V Panther tanks in a trench, and another showing a tank in a field. The page footer includes a list of other languages and a search bar.



Contexte

Introduction

- ❖ Présentation
- ❖ Programme
- ❖ Intérêt
- ❖ The german tank problem
- ❖ Problème du char d'assaut allemand

Contexte

Statistique exploratoire

Bases (et vocabulaire) de la statistique (exploratoire)

- la statistique a pour objectif d'étudier et d'aider à la compréhension de tous les phénomènes ou processus dans lesquels existe une certaine variabilité
 - discipline expérimentale (ou opérationnelle)
- les probabilités (le calcul des probabilités) ont pour objectif de formaliser les phénomènes ou processus dans lesquels intervient le hasard
 - discipline formelle (théorique ?), branche des mathématiques
- ces deux disciplines sont liés :
 - lien théorie ↔ pratique
 - lien hasard ↔ variabilité



Introduction

Statistique
exploratoire

- ❖ Étapes de la statistique exploratoire
- ❖ Démarche de la statistique exploratoire
- ❖ Intérêt de la statistique exploratoire
- ❖ C'est quoi la variabilité ?
- ❖ Mais non !!!

Bases (et
vocabulaire) de la
statistique
(exploratoire)

Statistique exploratoire



Étapes de la statistique exploratoire

Introduction

Statistique exploratoire

❖ Étapes de la statistique exploratoire

❖ Démarche de la statistique exploratoire

❖ Intérêt de la statistique exploratoire

❖ C'est quoi la variabilité ?

❖ Mais non !!!

Bases (et vocabulaire) de la statistique (exploratoire)

1. présentation de l'objet d'étude et du sujet d'étude (sur quoi ? pourquoi ?)
2. identification de la population ciblée et des variables étudiées (sur quoi ? comment ?)
3. présentation des données détaillées et validation
4. conception de l'analyse (comment analyser ?)
5. réalisation des traitements
6. interprétation statistique (que conclure ?)
7. restitution (comment présenter ?)



Démarche de la statistique exploratoire

Introduction

Statistique exploratoire

❖ Étapes de la statistique exploratoire

❖ Démarche de la statistique exploratoire

❖ Intérêt de la statistique exploratoire

❖ C'est quoi la variabilité ?

❖ Mais non !!!

Bases (et vocabulaire) de la statistique (exploratoire)

- la démarche du statisticien est celle
 - du médecin
 - du traducteur
 - de l'enquêteur



Intérêt de la statistique exploratoire

Introduction

Statistique exploratoire

- ❖ Étapes de la statistique exploratoire
- ❖ Démarche de la statistique exploratoire

❖ Intérêt de la statistique exploratoire

- ❖ C'est quoi la variabilité ?
- ❖ Mais non !!!

Bases (et vocabulaire) de la statistique (exploratoire)

- analyser la variabilité
- comprendre l'origine de la variabilité
- structurer la variabilité
- décider dans un contexte variable



C'est quoi la variabilité ?

Introduction

Statistique exploratoire

- ❖ Étapes de la statistique exploratoire
- ❖ Démarche de la statistique exploratoire
- ❖ Intérêt de la statistique exploratoire
- ❖ C'est quoi la variabilité ?
- ❖ Mais non !!!

Bases (et vocabulaire) de la statistique (exploratoire)





Mais non !!!

Introduction

Statistique
exploratoire

- ❖ Étapes de la statistique exploratoire
- ❖ Démarche de la statistique exploratoire
- ❖ Intérêt de la statistique exploratoire
- ❖ C'est quoi la variabilité ?
- ❖ Mais non !!!

Bases (et
vocabulaire) de la
statistique
(exploratoire)

La variabilité c'est

le nombre de personnes que vous avez vu sur
la photo !



Introduction

Statistique
exploratoire

Bases (et
vocabulaire) de la
statistique
(exploratoire)

- ❖ Objet et sujet d'étude
- ❖ Objet d'étude
- ❖ Moyens d'étude
- ❖ Types de variable
- ❖ Variables quantitatives
- ❖ Variables qualitatives
- ❖ Remarques
- ❖ Forme des données statistiques
- ❖ Exemple
- ❖ Méthode d'analyse

Bases (et vocabulaire) de la statistique (exploratoire)



Objet et sujet d'étude

Introduction

Statistique exploratoire

Bases (et vocabulaire) de la statistique (exploratoire)

❖ Objet et sujet d'étude

❖ Objet d'étude

❖ Moyens d'étude

❖ Types de variable

❖ Variables quantitatives

❖ Variables qualitatives

❖ Remarques

❖ Forme des données statistiques

❖ Exemple

❖ Méthode d'analyse

- ce que j'étudie : la photo de mariage
- ce que je cherche à savoir : combien de gens étaient sur la photo ?
- ce que j'étudie : un médicament - le baclofène, décontractant connu
- ce que je cherche à savoir : doit-on l'autoriser pour soigner l'alcoolisme ?



Objet d'étude

Introduction

Statistique exploratoire

Bases (et vocabulaire) de la statistique (exploratoire)

❖ Objet et sujet d'étude

❖ **Objet d'étude**

❖ Moyens d'étude

❖ Types de variable

❖ Variables quantitatives

❖ Variables qualitatives

❖ Remarques

❖ Forme des données statistiques

❖ Exemple

❖ Méthode d'analyse

- population : ce que j'étudie
- individus : les éléments de la population
- échantillon : un sous-ensemble de la population
- ★ comment l'ai je obtenu ?



Moyens d'étude

Introduction

Statistique
exploratoire

Bases (et
vocabulaire) de la
statistique
(exploratoire)

❖ Objet et sujet
d'étude

❖ Objet d'étude

❖ Moyens d'étude

❖ Types de variable

❖ Variables
quantitatives

❖ Variables
qualitatives

❖ Remarques

❖ Forme des
données statistiques

❖ Exemple

❖ Méthode
d'analyse

- on résume l'objet d'étude par des données pertinentes
 - les variables (symptômes du médecin, indices de l'enquêteur)
- ces variables doivent caractériser l'objet d'étude



Types de variable

Introduction

Statistique exploratoire

Bases (et vocabulaire) de la statistique (exploratoire)

❖ Objet et sujet d'étude

❖ Objet d'étude

❖ Moyens d'étude

❖ Types de variable

❖ Variables quantitatives

❖ Variables qualitatives

❖ Remarques

❖ Forme des données statistiques

❖ Exemple

❖ Méthode d'analyse

- les variables qui font référence à une quantité
 - variables quantitatives
- les variables qui font référence à une qualité
 - variables qualitatives
- des variables plus.... complexes
 - textuelles



Variables quantitatives

Introduction

Statistique exploratoire

Bases (et vocabulaire) de la statistique (exploratoire)

❖ Objet et sujet d'étude

❖ Objet d'étude

❖ Moyens d'étude

❖ Types de variable

❖ Variables quantitatives

❖ Variables qualitatives

❖ Remarques

❖ Forme des données statistiques

❖ Exemple

❖ Méthode d'analyse

- les variables que l'on mesure
 - variables continues (réelles ?, décimales ?)
- les variables que l'on compte
 - variables entières (discrètes ?)



Variables qualitatives

Introduction

Statistique exploratoire

Bases (et vocabulaire) de la statistique (exploratoire)

❖ Objet et sujet d'étude

❖ Objet d'étude

❖ Moyens d'étude

❖ Types de variable

❖ Variables quantitatives

❖ Variables qualitatives

❖ Remarques

❖ Forme des données statistiques

❖ Exemple

❖ Méthode d'analyse

- les variables sans ordre
 - variables nominales (catégorielles, non ordonnées)
- les variables avec ordre
 - variables ordinales (ordonnées, niveaux)



Remarques

Introduction

Statistique exploratoire

Bases (et vocabulaire) de la statistique (exploratoire)

- ❖ Objet et sujet d'étude
- ❖ Objet d'étude
- ❖ Moyens d'étude
- ❖ Types de variable
- ❖ Variables quantitatives
- ❖ Variables qualitatives
- ❖ Remarques
- ❖ Forme des données statistiques
- ❖ Exemple
- ❖ Méthode d'analyse

- ★ ne pas confondre type et codage
- ★ attention aux types mixtes



Forme des données statistiques

Introduction

Statistique exploratoire

Bases (et vocabulaire) de la statistique (exploratoire)

❖ Objet et sujet d'étude

❖ Objet d'étude

❖ Moyens d'étude

❖ Types de variable

❖ Variables quantitatives

❖ Variables qualitatives

❖ Remarques

❖ Forme des données statistiques

❖ Exemple

❖ Méthode d'analyse

- l'organisation générale du tableau est

Tableau individus \times variables

	<i>variables</i>					
	x_1^1	x_1^2	...	x_1^i	...	x_1^p
	x_2^1	x_2^2	...	x_2^i	...	x_2^p
	\vdots	\vdots		\vdots		\vdots
<i>individus</i>	x_r^1	x_r^2	...	x_r^i	...	x_r^p
	\vdots	\vdots		\vdots		\vdots
	x_n^1	x_n^2	...	x_n^i	...	x_n^p



Exemple

Introduction

Statistique exploratoire

Bases (et vocabulaire) de la statistique (exploratoire)

❖ Objet et sujet d'étude

❖ Objet d'étude

❖ Moyens d'étude

❖ Types de variable

❖ Variables quantitatives

❖ Variables qualitatives

❖ Remarques

❖ Forme des données statistiques

❖ Exemple

❖ Méthode d'analyse

	taille	sexe	nbrep
1	171	F	7
2	173	H	6
3	167	F	5
4	176	H	5
5	167	F	6
6	175	H	4
7	157	F	8
.	.	.	.
.	.	.	.
.	.	.	.
40	179	H	7
41	173	H	5
42	173	F	6
43	168	H	4
44	170	H	8
45	178	H	7



Méthode d'analyse

Introduction

Statistique exploratoire

Bases (et vocabulaire) de la statistique (exploratoire)

- ❖ Objet et sujet d'étude
- ❖ Objet d'étude
- ❖ Moyens d'étude
- ❖ Types de variable
- ❖ Variables quantitatives
- ❖ Variables qualitatives
- ❖ Remarques
- ❖ Forme des données statistiques
- ❖ Exemple
- ❖ Méthode d'analyse

- on travaille d'abord variable par variable
- statistique univariée
- puis on s'intéresse aux variables deux à deux
- pour étudier les liens
- statistique bivariée
- enfin on analyse le tableau globalement
- pour faire une synthèse
- statistique multivariée
- ★ pas étudiée dans ce cours