

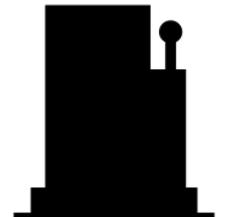
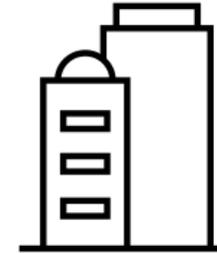
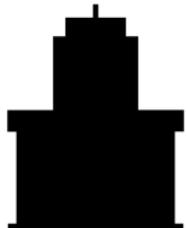
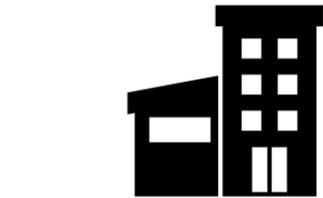
MINI-INTRODUCTION EN STATISTIQUE

Purpose and Sustainability
Purpose Track - Mémoire

18 septembre 2020 - Chang-Wa HUYNH

Rappel : articulation des quatre méthodes proposées

Hypothèse de recherche : X \longrightarrow Y
(e.g. Age, Genre) (e.g. Légitimité du leader)



Sondages
(validité externe, corrélation)

Que sont les statistiques ?

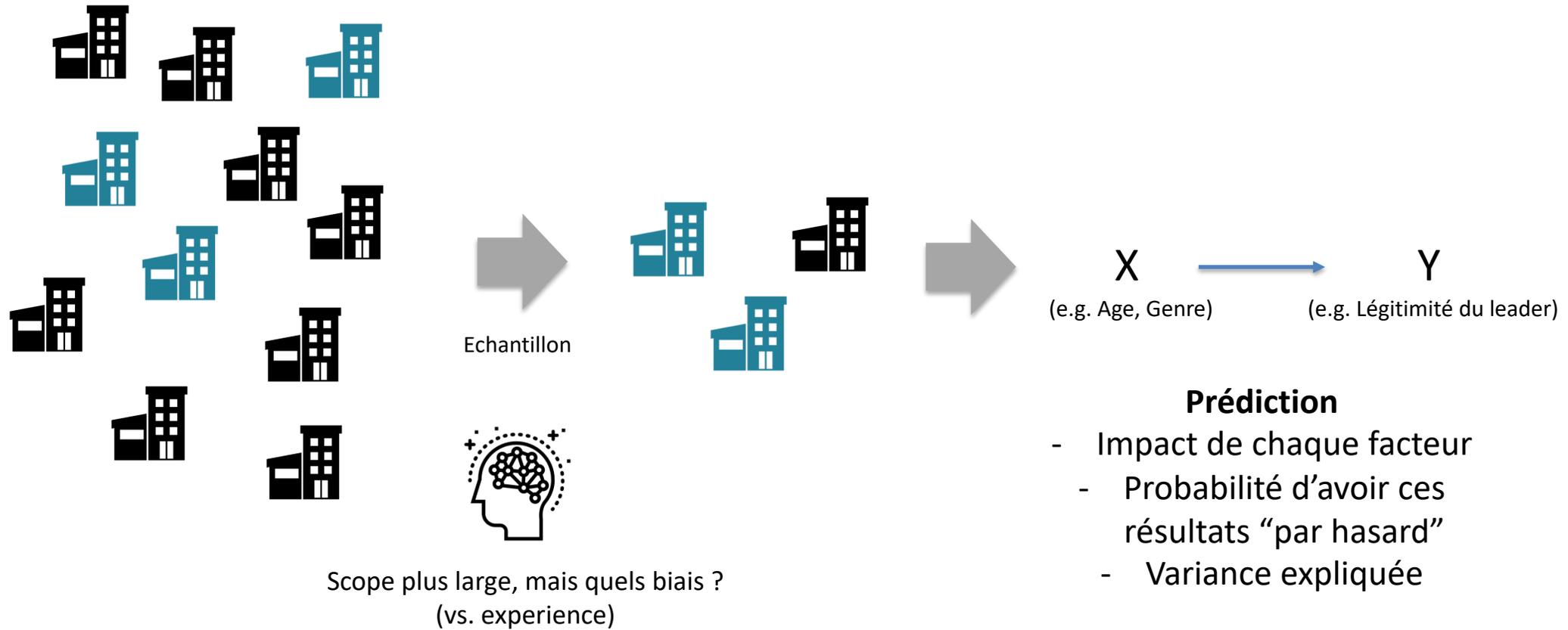
Probabilités ↔ Statistiques

Modèles connus / postulés
Simulations
Modèles épidémiologiques
Modèles financiers
Gestion de files d'attente

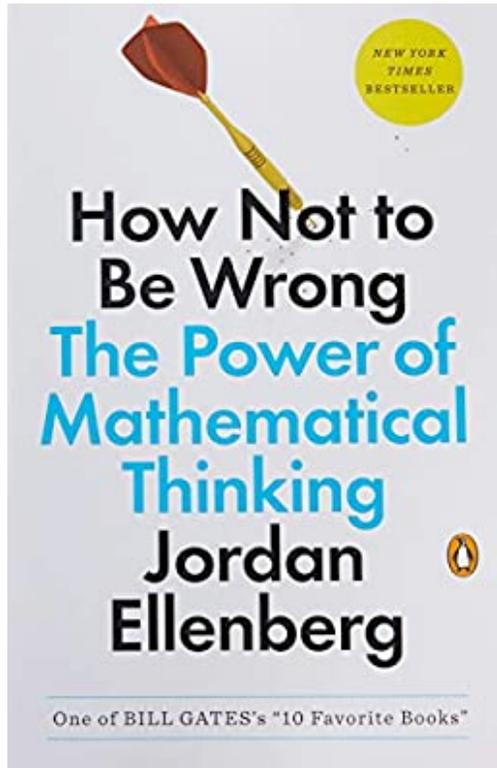
Etudes de données existantes
Statistiques descriptives
Statistiques mathématiques
Tests d'hypothèses
Prédiction
...

Big Data
(+ informatique)

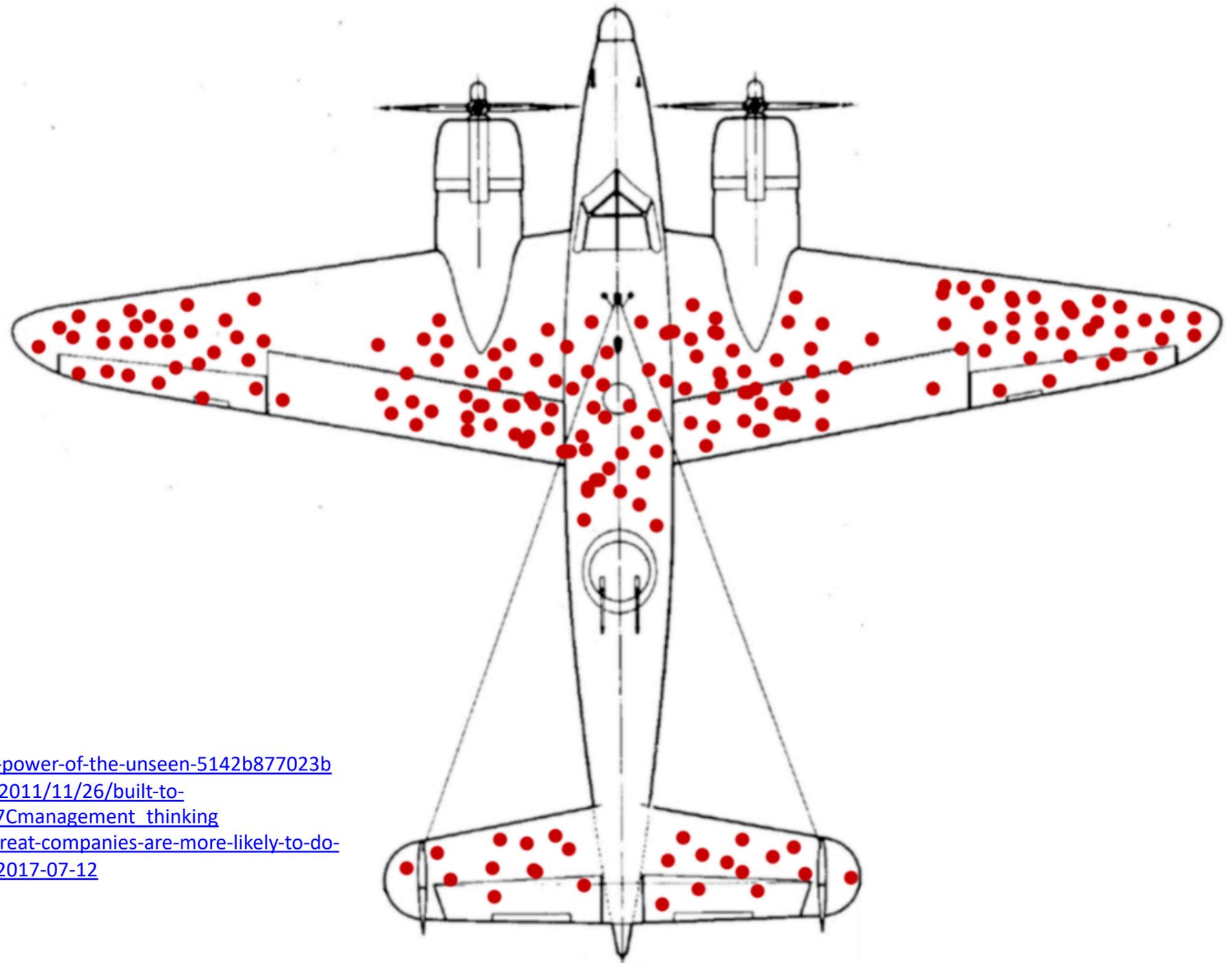
Analyse des sondages



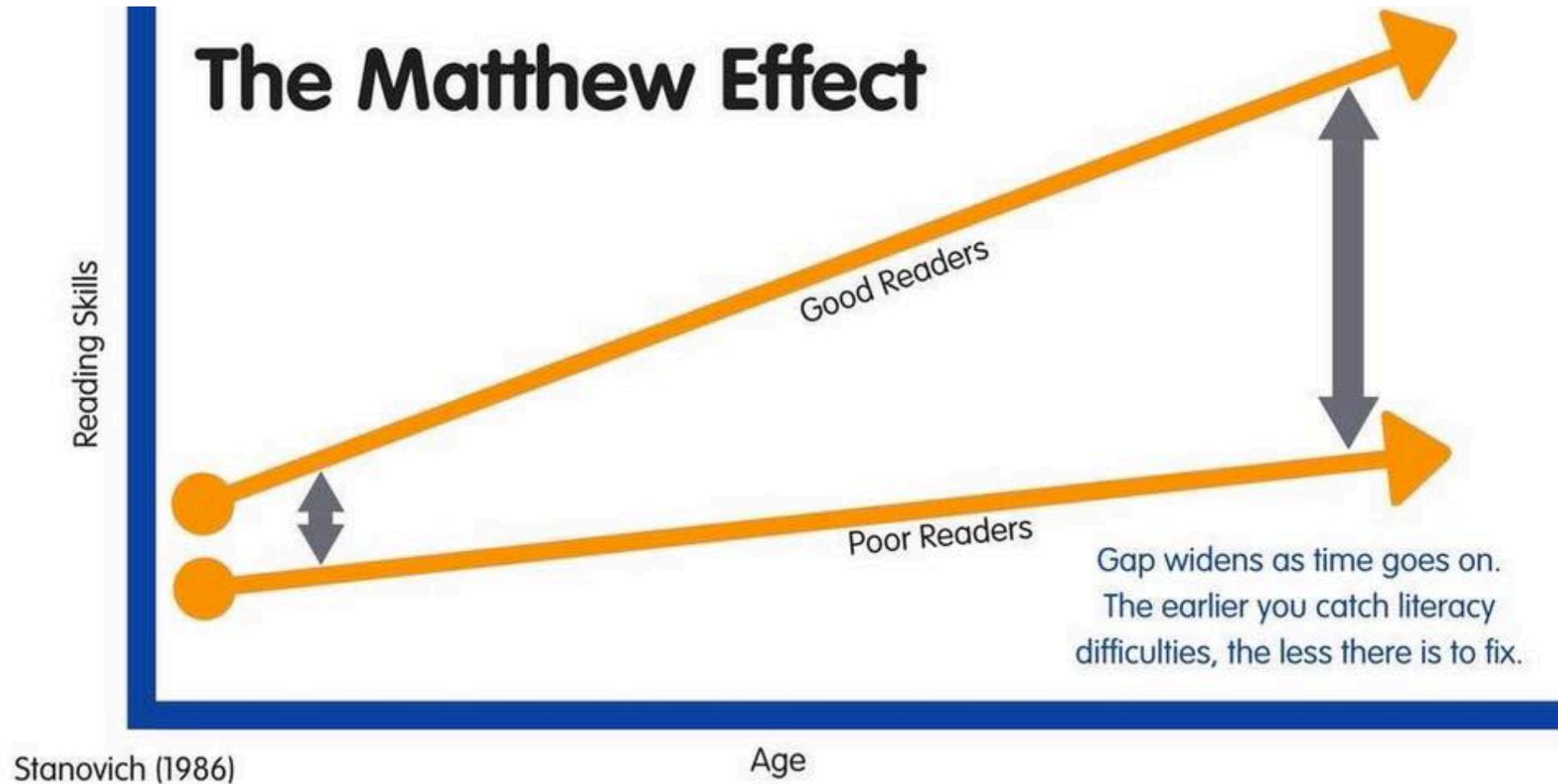
Biais du survivant



- <https://medium.com/@zandercutt/the-power-of-the-unseen-5142b877023b>
- [https://www.economist.com/business/2011/11/26/built-to-last?fsrc=nlw%7Cmgt%7C11-30-2011%7Cmanagement thinking](https://www.economist.com/business/2011/11/26/built-to-last?fsrc=nlw%7Cmgt%7C11-30-2011%7Cmanagement%20thinking)
- <https://www.marketwatch.com/story/great-companies-are-more-likely-to-do-really-badly-over-time-than-really-well-2017-07-12>



Simultanéité / causalité inversée

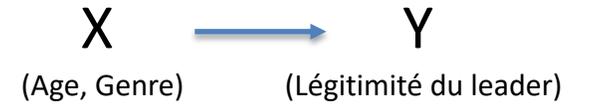


<http://www.garfield.library.upenn.edu/merton/matthew1.pdf>

https://en.wikipedia.org/wiki/Matthew_effect

Un exemple

Rappel
Hypothèse de recherche



Participant	Genre	Age	Approbation (/10)
1	F	23	1
2	F	21	3
3	F	20	4
4	F	32	6
5	F	30	8
6	F	35	10
7	H	20	5
8	H	32	8
-	H	21	2
-	F	27	3

Moyenne d'approbation : 5.6/10
75% de femmes et 25 ans d'âge moyen

Moyenne d'approbation chez les femmes : 5.3/10
Moyenne d'approbation chez les hommes: 6.5/10

Tendance plus difficile à dégager pour l'âge

Non observés
Pourquoi ? Impact ?

En pratique

Résultats sous SPSS



Autres logiciels



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.849 ^a	.720	.609	1.85991

a. Predictors: (Constant), Genre, Age

72% de la variance est expliquée par les facteurs utilisés

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	44.579	2	22.289	6.443	.041 ^b
	Residual	17.296	5	3.459		
	Total	61.875	7			

a. Dependent Variable: Approbation

b. Predictors: (Constant), Genre, Age

Le modèle global est significatif (<0.05)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3.804	3.219		-1.182	.290
	Age	.396	.113	.831	3.507	.017
	Genre	-1.497	1.522	-.233	-.984	.370

a. Dependent Variable: Approbation

Significatif (<0.05), chaque année supplémentaire augmente le score d'approbation de 0.4 points

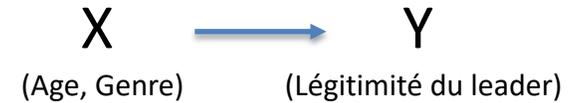
Non significatif (mais ne veut pas dire que cet effet n'existe pas)

Aurions-nous pu obtenir ces données "par hasard" ? Plus l'effet est subtil, plus il faut de participants pour le détecter

Un (nouvel) exemple avec un effet moins marqué

Rappel

Hypothèse de recherche



Participant	Genre	Age	Ancienne Approbation (/10)	Nouvelle Approbation (/10)
1	F	23	1	5
2	F	21	3	6
3	F	20	4	5.5
4	F	32	6	6.5
5	F	30	8	6.5
6	F	35	10	6
7	H	20	5	5.5
8	H	32	8	6
-	H	21	2	5
-	F	27	3	5

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.657 ^a	.432	.204	.46172

a. Predictors: (Constant), Genre, Age

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.809	2	.405	1.898	.244 ^b
	Residual	1.066	5	.213		
	Total	1.875	7			

a. Dependent Variable: Approbation

b. Predictors: (Constant), Genre, Age

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.366	.799		5.463	.003
	Age	.053	.028	.641	1.897	.116
	Genre	.122	.378	.109	.324	.759

a. Dependent Variable: Approbation

Quelques étapes

- **Quels sont les biais de mon échantillon ?**
- **Quelles sont les variables explicatives intéressantes pour mon hypothèse de recherche ?**
 - Limiter le nombre de variables. Celles-ci ne doivent pas être trop corrélées entre elles.
- **Importer les données dans SPSS: File/Import/Excel**
- **Vérifier le type (en bas de la fenêtre de données/Variable view):**
 - Scale (nombre) / Nominal (catégories, comme le genre, l'industrie, etc.)
- **Statistiques descriptives: Analyze/Descriptive statistics/Descriptives**
- **Modèle linéaire: Analyze/Regression/Linear**
 - Hypothèses du modèle et quick start: <https://statistics.laerd.com/spss-tutorials/linear-regression-using-spss-statistics.php>
 - Trois questions : le modèle global est-il significatif ? Quelle variance explique t-il ? Quels sont les facteurs explicatifs intéressants (significatif ? magnitude ? signe ?)
- **Concentrer vos analyses sur le modèle le plus intéressant**

Annexe: recodage

- **Les fichiers excel que vous exportez (e.g. Google Forms) peuvent contenir des données numériques mais sous forme de texte.**
 - “pas d’accord”, “plutôt d’accord”, etc.
- **Il est possible de les recoder sous forme numérique dans Excel pour pouvoir exploiter les données.**
 - Transformer “pas d’accord” en -1, “Plutôt pas d’accord” en -0.5, etc.
- **La fonction à utiliser est « RECHERCHEV » ou « VLOOKUP » (pour les Excel en français et anglais respectivement).**
 - La syntaxe de la fonction est =RECHERCHEV(Valeur dans le sondage; Zone de la table de recoding (en commençant par la colonne où se trouve les valeurs dans le sondage); Numéro de la colonne où se trouve la valeur numérique associée (2 si c’est à côté); faux (pour indiquer qu’il ne faut considérer que les matching exacts))
 - Pour un exemple : <https://sites.google.com/a/uci.edu/ss3a/advanced-excel-functions> (section « B. VLOOKUP and Recoding »)
 - Les tables de recoding doivent être sur une feuille Excel séparée (et non sur le côté, comme sur le tutorial) pour ne pas être importées dans SPSS