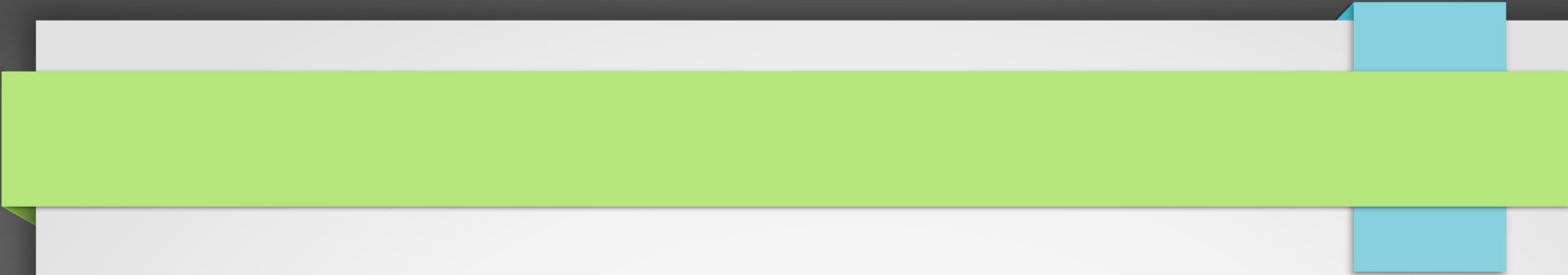


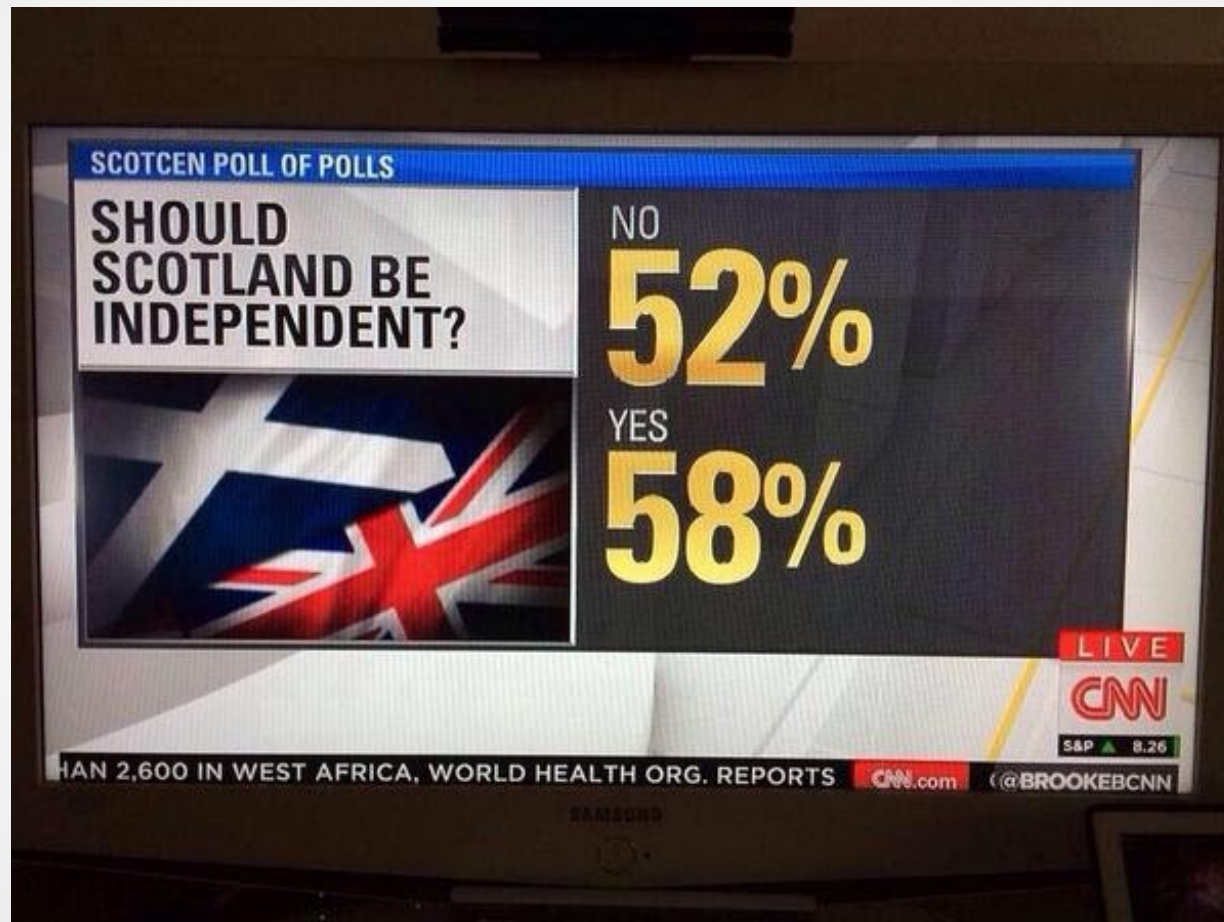
Statistique quantitative

Dr J.C. Bartier
Pôle Anesthésie-Réanimation-SAMU-CESU
Hôpitaux Universitaires de Strasbourg



S'il ne me restait plus qu'une heure à vivre,
j'aimerais la passer dans un cours de statistiques :
elle me semblerait tellement plus longue...

Tout le monde utilise les statistiques...




3 objectifs

- Comprendre la démarche quantitative à travers l'analyse d'un article
- Comprendre l'importance de la saisie et de l'organisation de la saisie des données
- Aborder l'analyse statistique de base

L'analyse statistique quantitative se nourrit de l'analyse statistique qualitative ... et réciproquement

RESEARCH

CHRISTMAS 2013: RESEARCH

**The survival time of chocolates on hospital wards:
covert observational study** OPEN ACCESS

Parag R Gajendragadkar *cardiology specialist registrar*¹, Daniel J Moualel *ENT surgery specialist registrar*², Phillip L R Nicolson *haematology specialist registrar*³, Felicia D Adjei *core medical trainee*¹, Holly E Cakebread *foundation year doctor*¹, Rudolf M Duehmke *cardiology specialist registrar*¹, Claire A Martin *cardiology specialist registrar*¹

¹Department of Cardiology, Bedford Hospital, Bedford MK42 8DJ, UK; ²Department of Ear, Nose and Throat Surgery, Great Western Hospital, Swindon, UK; ³Department of Haematology, University Hospital of North Staffordshire, Stoke-on-Trent, UK

Anatomie d'un article

- Introduction
- Matériel et méthode
- Résultat
- Discussion
- Conclusion
- Bibliographie

Quelle est la question ?

- Toujours poser la « question de recherche »
 - Unique et précise : Résumer un travail par une question simple
 - Pourquoi ? (pourquoi s'intéresser à cette question)
- Exemple
 - Thème : L'obésité a-elle une influence sur la réussite d'une analgésie péridurale ?
 - Question : existe t'il une relation entre un IMC élevé et l'échec d'une analgésie péridurale ?
 - Titre de l'article / mémoire
 - Plus difficile que cela en a l'air...

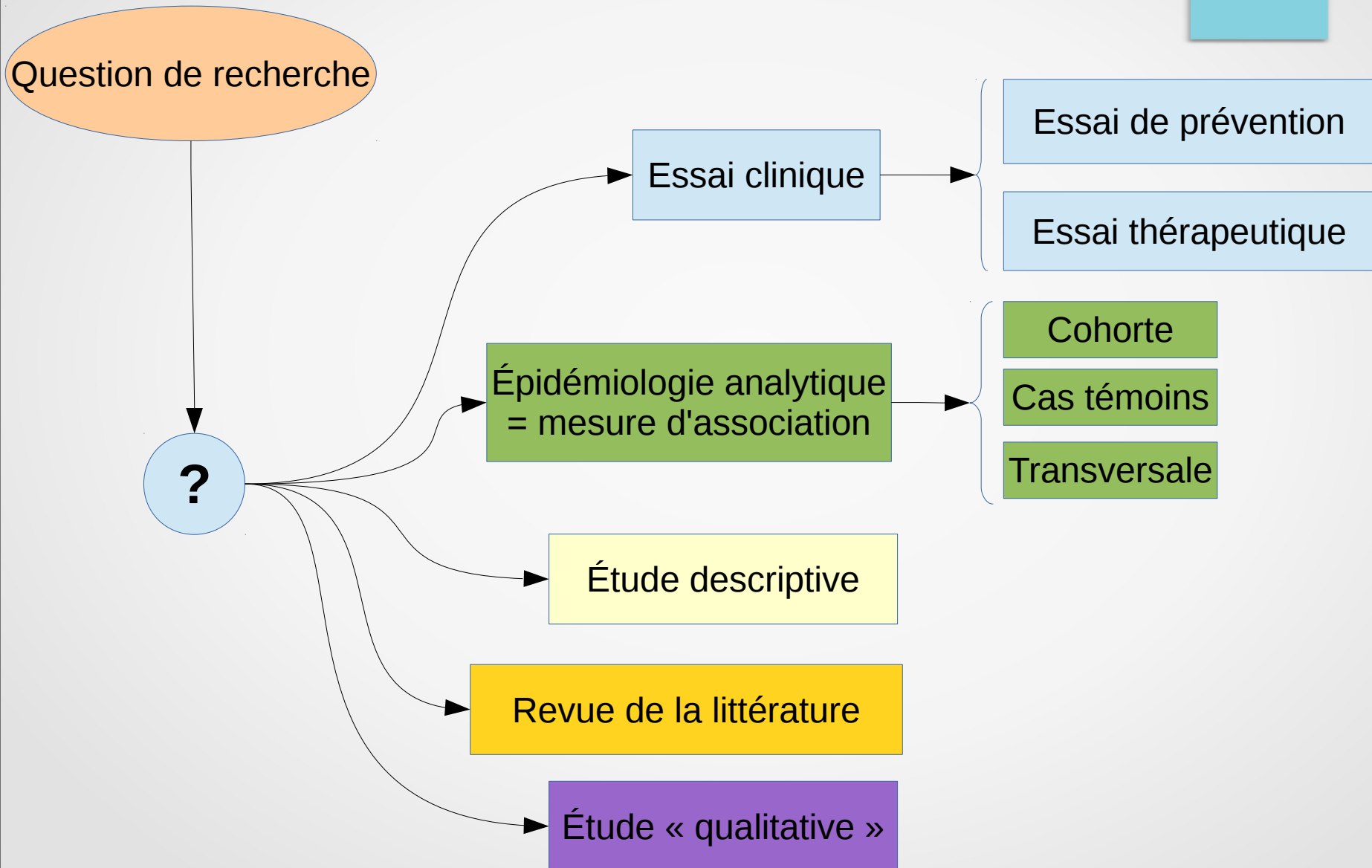
CONTEXTE ET JUSTIFICATION

- L'augmentation de la consommation de chocolat a été associée à une réduction des risques de développement de troubles cardio-métabolique [1].
- En outre, la variation nationale dans la consommation de chocolat est positivement corrélée avec le nombre de prix Nobel par unité de population [2].
- La consommation du chocolat en milieu hospitalier est un phénomène relativement courant et les cadeaux de patients et des familles représentent une grande proportion de la consommation de chocolat des travailleurs de la santé.
- Les preuves anecdotiques suggèrent que différentes variétés de chocolat pré-emballés paraissent durer plus ou moins longtemps, et que la voracité de certains professionnels de santé semble être une source fréquente de conflits au travail.

Que veut-on prouver ?

- Les objectifs de l'étude sont d'essayer d'identifier les variations de la consommation de chocolat parmi les professionnels de santé en milieu hospitalier en regardant:
 1. Le temps de survie d'un chocolat dans le contexte d'un service
 2. Un modèle mathématique qui explique la consommation de chocolat observé
 3. Les différences de temps de survie selon les marques de chocolat
 4. Les différences de consommation de chocolat parmi les professionnels de santé stratifiés par type d'emploi.

Quelle méthodologie ?



Type de protocole

- Le **temps de survie de chocolats** dans un service hospitalier:
- une étude observationnelle secrète
- Objectif: quantifier la consommation de chocolats dans un service hospitalier.
- Etude **multicentrique, prospective**, de type **observationnelle** secrète.
- Observation de **quatre services** dans **trois hôpitaux** (où les auteurs ont travaillé) dans le Royaume-Uni.
- Sujets d'étude: Boîtes de **Quality Street** (Nestlé) et **Roses** (Cadbury) posées sur le comptoir du service de soins et accessible à tout le monde.

On compare 2 traitements

Des boîtes Standard de 350 g

- **Roses** fabriqués par Cadbury, Royaume-Uni
- **Quality Street** fabriqués par Nestlé, Suisse seront utilisés.



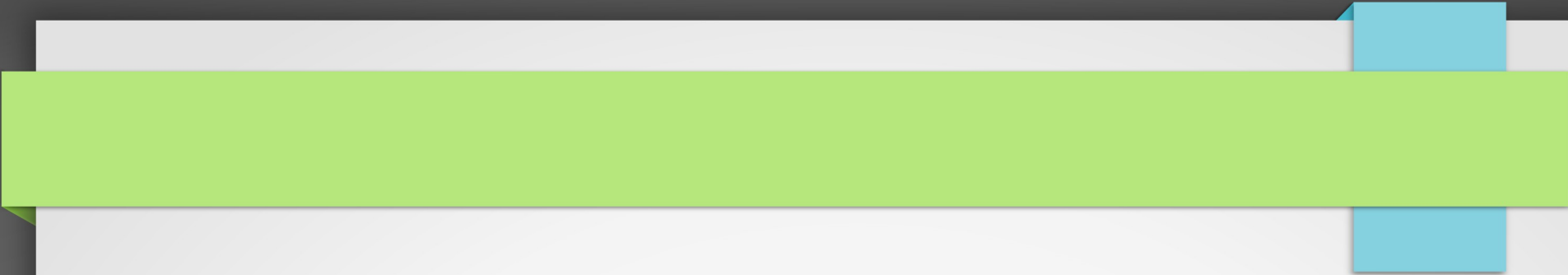
Les observateurs placent secrètement deux boîtes de 350 g de chocolats, Quality Street et Roses dans chaque service (huit boîtes ont été utilisées dans l'étude contenant un total de 258 chocolats). Ces boîtes ont été gardés sous surveillance discrète et continue, avec une mesure du temps à chaque consommation d'un chocolat.

Le protocole

- Les boîtes de chocolat seront placés dans chaque service sans explications ni avertissements entre 0900-1000.
- Elles seront placées côte à côte à un endroit stratégique déterminé par l'observateur.
- Les boîtes resteront sous surveillance visuelle continue par des observateurs qui connaissent bien les personnels du service et leurs fonctions.
- L'observateur ne devra pas montrer aux personnel présents qu'il observe les chocolats (observationnel secret). L'observateur enregistre les données sur le formulaire de collecte des données joint (Annexe 1) jusqu'à ce que les deux boîtes de chocolat soient vidées.
- L'observateur ne participe pas aux soins des patients pendant la période d'observation.

Aspects éthiques

- Les coûts des boîtes ont été calculés et une part égale payée individuellement par tous les auteurs de l'étude. **Aucun fond externe** n'a été nécessaire.
- Tous les événements indésirables et indésirables graves seront rapportés. En dehors de **réactions allergiques** nous ne prévoyons pas d'événements indésirables.
- Nous ne avons pas cherché **l'approbation du comité éthique** pour l'étude. Aucune données individuelles identifiables ne seront recueillies.
- Nous avons pris en considération le **risque de gain de poids** liés à la mise à disposition de chocolat supplémentaire chez des sujets vulnérables.
- Compte tenu de la **mise à disposition gratuite** des produits et le fait que **l'absence de coercition**, nous jugeons ce problème mineur.
- En outre, les quantités de chocolat susceptibles d'être consommés par une seule personne cette étude sont faibles. Il y a un petit risque de réactions allergiques, mais nous avons estimés que compte tenu de la mise à disposition gratuite de ces marques dans tout l'hôpital, il parait peu probable que ce soit un risque important et méconnu des consommateurs.
- l'étude sera effectué en **conformité avec les recommandations pour les médecins impliqués dans la recherche sur des sujets humains adoptés par la 18e Assemblée Médicale Mondiale, Helsinki 1964** et révisions ultérieures.

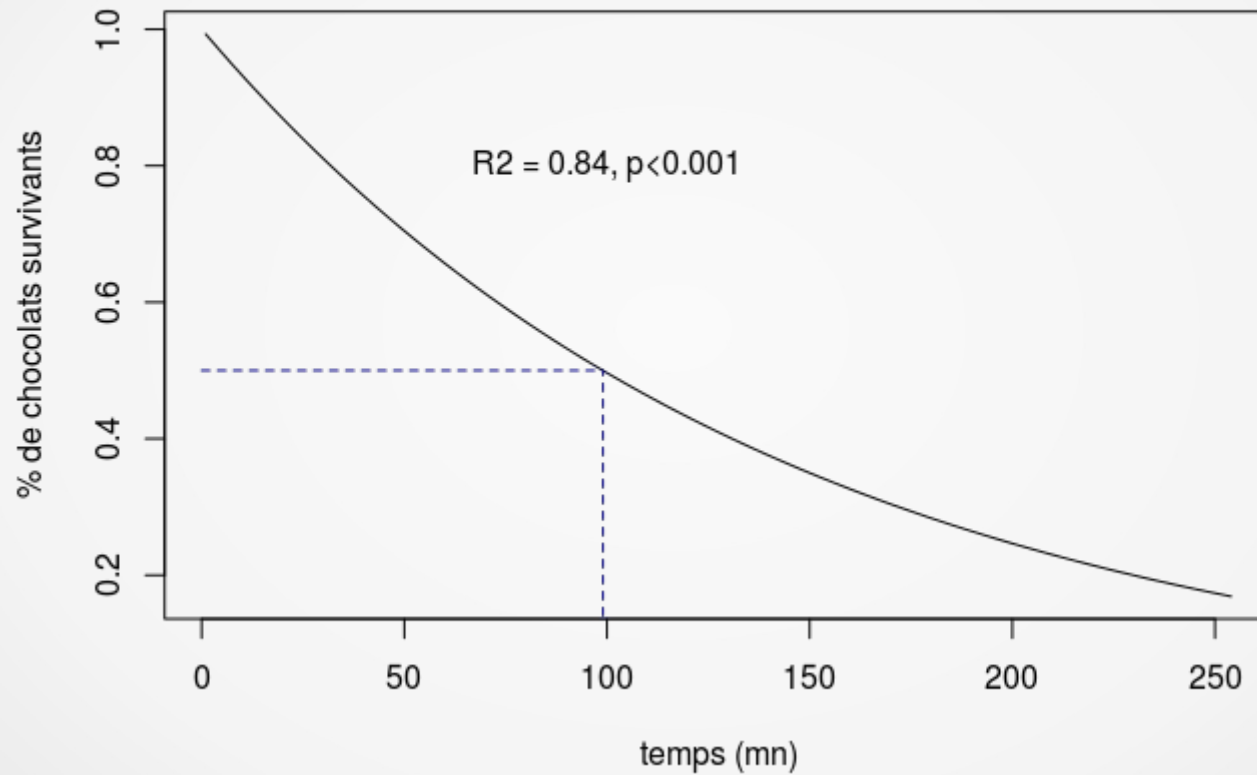
- 
- **critère(s) de jugement**
 - temps de survie **médian** d'un chocolat.

Résultats

- 191 sur 258 (**74%**) chocolats ont été mangés.
- La durée totale moyenne d'observation a été de **254 minutes** (intervalle de confiance à 95 % de 179 à 329).
- La durée **médiane de survie** d'un chocolat était de **51 minutes** (39-63).
- Le modèle de la consommation de chocolat était **non linéaire**,
 - avec un taux initial rapide de la consommation
 - qui a ralenti avec le temps.
 - Un modèle de décroissance exponentielle explique ces résultats
 - (modèle $R^2 = 0,844$, $P < 0,001$),
- avec une **demi-vie** de survie (temps nécessaire pour que 50% des chocolats soient consommés) de **99 minutes**.

Modèle exponentiel

Temps de survie des chocolats



$$Y = e^{-0,007 \cdot t}$$

Résultats secondaires

- Le temps moyen pour ouvrir une boîte de chocolats était de **12 minutes** (95% d'intervalle de confiance 0-24).
- Les Chocolats 'Quality Street' ont survécu plus longtemps que les chocolats 'Roses' chocolats
 - (hazard ratio pour la survie de Roses v Quality Street 0,70, intervalle de confiance à 95% 0.53 à 0.93, **P = 0,014**).
- Les pourcentages les plus élevés de chocolats ont été consommés par
 - des AS (28%) et les infirmières (28%),
 - suivis par les médecins (15%).

En conclusion

- Ce qui est déjà connu sur le sujet
 - **Peu de données quantitatives sont disponibles sur la survie du chocolat** dans un environnement hospitalier et sur la consommation de chocolat par les professionnels de santé, bien que des preuves anecdotiques soient abondantes.
- Ce que cette étude ajoute
 - Cette étude suggère que le **temps de survie d'un morceau de chocolat est court**, moins d'une heure, et que le taux initial de consommation d'une boîte de chocolat boîte est rapide, mais ralentit avec le temps
 - Un **modèle de décroissance exponentielle** décrit le mieux cette tendance
 - Dans l'ensemble, les pourcentages les plus élevés de chocolats ont été consommés par les AS et les infirmières, suivie par les médecins

Donnée (Data)

- Ce que l'on observe, mesure
- Matière première du statisticien
- Tout travail scientifique génère et/ou manipule des données (data : Point sur lequel on fonde un raisonnement)
- 80 % du temps consacré à un travail scientifique est consacré à la préparation des données
- Le traitement statistique des données est le principal frein à une production rapide de résultats
 - Mauvaise présentation des données
 - Difficultés de compréhension entre le producteur de données et le statisticien

Données (Data)

- 2 grands types
 - Quantitatives (continues)
 - Qualitatives (on ne peut pas les additionner) ou discrètes
 - Catégories non recouvrantes
 - Habituellement moins de 100 catégories
 - Ex : sexe (Homme, Femme, inconnu)
 - On peut assez facilement transformer une variable quantitative en qualitative (discrétisation). L'inverse n'est pas vrai :
 - EX : age et tranche d'age (INSEE)
- La distinction est importante dans le choix des méthodes statistiques applicables.

Variables qualitatives

- Elles servent à faire des groupes (sexe, CSP)
- Plusieurs type
 - Nominale : simple énumération, l'ordre n'a pas d'importance : Homme/Femme
 - Ordinale : l'ordre à un sens (mais on ne sait pas dire de combien un niveau est différent d'un autre : item de Likert, Score de Glasgow
 - Echelles : variables ordinales dont l'espacement des niveaux est régulier : échelle de température
- Attention au var.qualitative qui se présentent comme des var. Quantitatives : on peut calculer un score de Glasgow moyen mais cela n'a pas de sens.

Les ensembles de données

- ensemble de données, fichier, feuille de calcul, tableur = dataset
- Format rectangulaire
 - en langage mathématique : ***matrice***
 - en terme de base données : ***table***
 - en langage de tableur : ***feuille de calcul***
- Un dataset est habituellement un tableau
 - constitué de **lignes** et de **colonnes**.
 - Une ligne contient une **observation (une personne)**,
 - Une colonne correspond à une **variable (ce que je mesure)**.

Anatomie d'un tableur

- Lignes
 - Première ligne = en-tête (**Header**) = nom colonne
 - 1 ligne = 1 unité expérimentale (observation)
- Colonnes
 - 1 colonne = 1 variable (age, profession , score...)
 - 1ere colonne = **identifiant**

	A	B	C	D	E	F	G
1	titre						
2	Objectif(s)						
3	Investigateur						
4	Institution	CESU 67					
5							
6	unité	AS	Date de	DEC	Note	Pre	Post
7	expérimental	IDE	naissance	TRAD	globale	test	test
8		AMB					
9							
10	ID	METIER	AGE	GRUPE	NOTE	PT	OT
11	1	DE	22	DEC	10	6	6
12	2	DE	25	DEC	8	5	6
13	3	DE	23	DEC	7	5	5
14	4	DE	31	DEC	8	4	5
15	5	S	20	DEC	7	3	5
16	6	S	21	TRAD	9	5	8
17	7	S	24	TRAD	10	8	8
18	8	S	25	TRAD	10	7	7
19	9	MB	28	TRAD	5	5	4
20	10	MB	30	TRAD	6	4	4
21							

Anatomie d'un tableur (2)

- Un certain nombre de règles sont à respecter pour que les données soient exploitables.
- La feuille de saisie est divisée en 3 zones horizontales:
 - la première, facultative, sert à stocker les métadonnées [<http://fr.wikipedia.org/wiki/M%C3%A9tadonn%C3%A9e>]
 - la seconde stocke le nom des colonnes
 - la troisième contient les données proprement dites
- La troisième zone est divisée verticalement en 2:
 - les colonnes les plus à gauche contiennent les variables de type "facteur" (factorielles : profession, type de pédagogie)
 - les colonnes les plus à droite contiennent les mesures

Métadonnées

- Métadonnées
 - ce sont des données à propos des données.
 - Elles servent à comprendre le contexte dans lequel se fait l'étude:
 - nom de l'auteur, des investigateurs
 - version du questionnaire
 - méthodes de description des variables: 'pour la rubrique sexe préciser H ou F'.
 - Cette zone est facultative ou faire l'objet d'un document séparé

Nom des colonnes

- une seule ligne.
- Sert à stocker le nom opérationnel des colonnes.
- Règles:
 - le nom est succinct (5 caractères)
 - il ne doit pas contenir d'espace ou de caractères qui pourrait être mal interprété (, ; / \ @ &)
 - remplacer les espace par des underscore (caractère 8) :
Ceci_est_un_exemple
 - pas de caractères accentués
 - Encodage : UTF-8
 - Ne pas fusionner les cellules

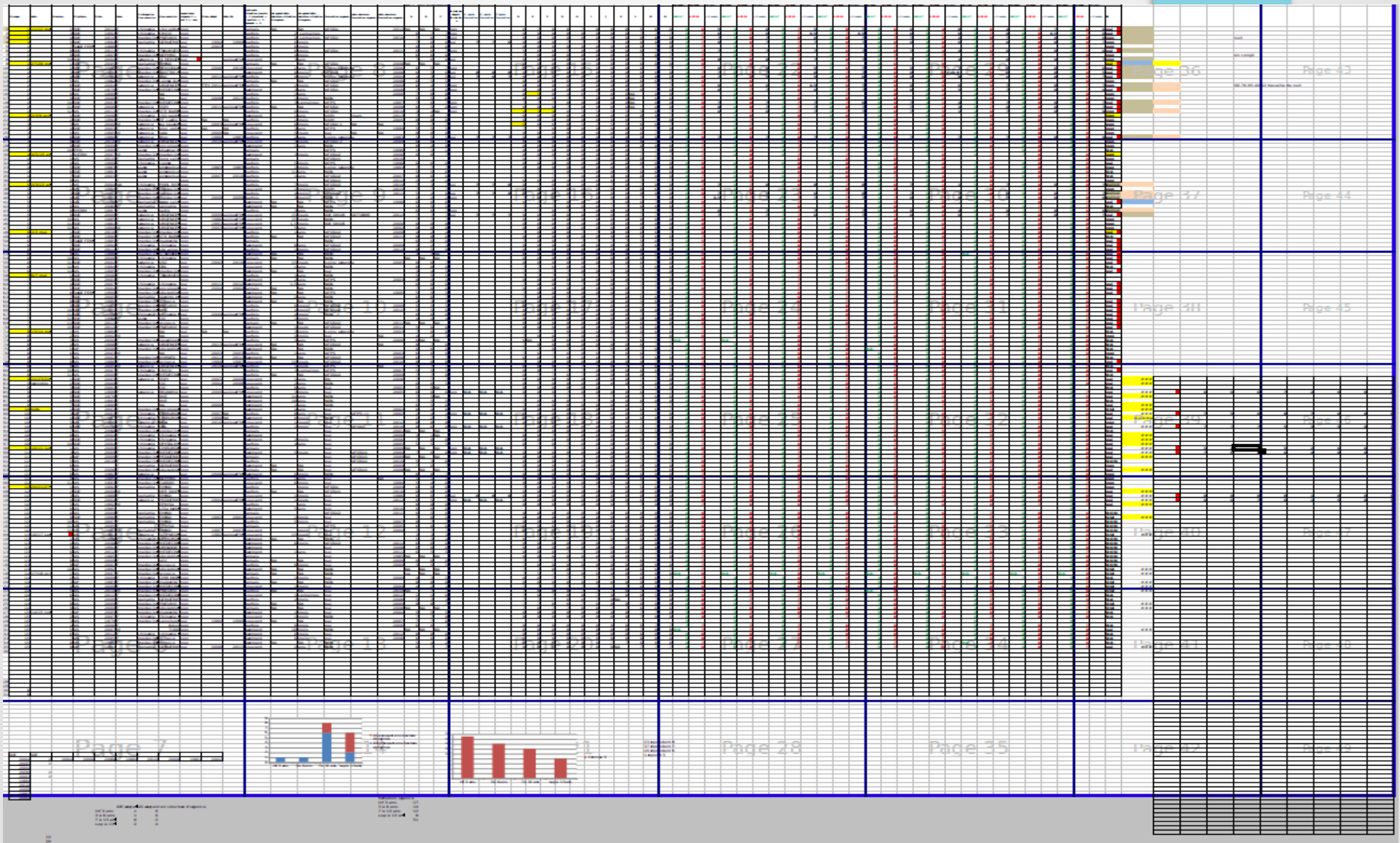
Mesures

- toujours utiliser le point décimal (jamais la virgule)
- une colonne ne peut contenir que des chiffres ou du texte, jamais des deux.
- Les dates de préférence au format ISO :
 - AAAA-MM-JJ ex. 2014-02-06
 - AAAA-MM-JJ HH:MM:SS ex. 2014-02-06 10:25:36
- Un logiciel ne sait pas distinguer les codes de couleur
- Un logiciel gère très bien les lettres
 - Non : 1, 1, 2, 2, 2, 0, 1, 2
 - Oui : H, H, F, F, F, NA, H, F
- Ne pas mélanger les majuscules et les minuscules
 - Homme <> Homme <> HOMME
- Valeur manquante
 - La déclarer explicitement (ne pas laisser de blanc)
 - Ne pas utiliser le zéro ou une valeur négative
 - Valeur par défaut : NA (not available).

Présentation des données

- 80 % du temps de travail d'un statisticien est consacré au nettoyage des données
- Tableur (Excel, Libre office, etc.)
 - Très bon moyen de saisie des données
 - Très mauvais outil de traitement des données
- Logiciels de traitement (R, SPSS, STATA,...)
 - Mauvais moyens de saisie
 - Bon moyen de traitement
- Bons outils = tableur + logiciel spécialisé
 - Recherche reproductible

Ce qu'il ne faut pas faire...



Une philosophie : recherche reproductible (reproducible research)

Le format CSV

- CSV = comma separated virgule (données séparées par une virgule)
- Format universel (fichier/enregistrer sous)
- Les colonne du tableur sont remplacées par des virgules
- Le nom du fichier se termine par **.csv**
- Il existe des variantes (tab,;)
- [\[http://fr.wikipedia.org/wiki/Comma-separated_values\]](http://fr.wikipedia.org/wiki/Comma-separated_values)

Format csv

Tableau

hopitaux.csv - LibreOffice Calc

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	calendrier	Sel	Col	Odi	Wis	Geb	Hus	3Fr	Mul	Hag	Dia	Alk	Sav	moyenne
2	2013-01-01	111	208	84	32	45	126	59	1	131	88	1	1	73.9166666666667
3	2013-01-02	80	197	69	49	42	125	38	1	112	89	1	1	67
4	2013-01-03	78	160	55	35	42	121	39	1	83	73	1	1	57.4166666666667
5	2013-01-04	65	170	67	24	30	121	42	1	92	93	1	1	58.9166666666667
6	2013-01-05	85	150	70	38	44	102	46	1	100	87	1	1	60.4166666666667
7	2013-01-06	68	167	79	36	43	93	38	1	90	77	1	1	57.8333333333333
8	2013-01-07	84	171	57	31	34	119	37	180	84	79	1	1	73.1666666666667
9	2013-01-08	84	168	69	23	44	113	42	1	82	69	1	1	58.0833333333333
10	2013-01-09	56	185	61	30	43	96	36	1	91	85	1	1	57.1666666666667
11	2013-01-10	75	184	64	37	45	77	44	1	81	66	1	1	56.3333333333333
12	2013-01-11	71	168	56	26	24	128	32	1	67	70	1	1	53.75
13	2013-01-12	64	163	72	31	33	116	34	1	95	75	1	1	57.1666666666667
14	2013-01-13	75	149	67	29	41	76	53	1	97	71	1	1	55.0833333333333
15	2013-01-14	64	173	63	35	39	123	37	168	103	86	1	1	74.4166666666667
16	2013-01-15	89	151	49	33	48	113	36	157	93	81	1	1	71
17	2013-01-16	63	160	57	26	29	117	50	149	98	68	1	1	68.25
18	2013-01-17	71	187	54	22	38	103	41	153	78	66	1	1	67.9166666666667
19	2013-01-18	78	184	71	25	33	105	43	81	93	73	1	1	65.6666666666667
20	2013-01-19	88	159	58	26	37	69	47	178	103	82	1	1	70.75

.CSV

hopitaux.csv x

```
1 calendrier,Sel,Col,Odi,Wis,Geb,Hus,3Fr,Mul,Hag,Dia,Alk,Sav,moyenne
2 2013-01-01,111,208,84,32,45,126,59,1,131,88,1,1,73.9166666666667
3 2013-01-02,80,197,69,49,42,125,38,1,112,89,1,1,67
4 2013-01-03,78,160,55,35,42,121,39,1,83,73,1,1,57.4166666666667
5 2013-01-04,65,170,67,24,30,121,42,1,92,93,1,1,58.9166666666667
6 2013-01-05,85,150,70,38,44,102,46,1,100,87,1,1,60.4166666666667
7 2013-01-06,68,167,79,36,43,93,38,1,90,77,1,1,57.8333333333333
8 2013-01-07,84,171,57,31,34,119,37,180,84,79,1,1,73.1666666666667
9 2013-01-08,84,168,69,23,44,113,42,1,82,69,1,1,58.0833333333333
10 2013-01-09,56,185,61,30,43,96,36,1,91,85,1,1,57.1666666666667
11 2013-01-10,75,184,64,37,45,77,44,1,81,66,1,1,56.3333333333333
12 2013-01-11,71,168,56,26,24,128,32,1,67,70,1,1,53.75
13 2013-01-12,64,163,72,31,33,116,34,1,95,75,1,1,57.1666666666667
14 2013-01-13,75,149,67,29,41,76,53,1,97,71,1,1,55.0833333333333
```

Organiser son travail

- Tableur
 - Données brutes (toujours conserver les originaux)
 - Données nettoyées (toujours travailler sur des copies)
 - Choisir son style
 - Solitaire : Open Office (Excel)
 - Collaboratif : cloud (Google drive) attention à la confidentialité...
- Notebook
 - Livre de bord horodaté
 - Métadonnées (dictionnaire des variables)
 - Où j'en suis
 - To Do (ce qu'il reste à faire)
- Dossier DOC
 - Articles
 - Biblio
- Dossier mémoire/thèse/article

4 éléments sont nécessaires pour le statisticien

- 1. Les données brutes.
- 2. Les données nettoyées (tidy)
- 3. Un lexique décrivant chaque variable et ses valeurs possibles.
- 4. La recette utilisée pour passer du point 1 aux points 2 et 3

Contacter le statisticien au stade du projet, jamais à la fin du travail...

En résumé

- Données brutes (raw data)
 - Feuille de papier
- Données non préparées (messy data)
 - Données transcrites dans un tableur
 - les entête de colonne contiennent des chiffres au lieu de lettres
 - des variables multiples stockées dans la même colonne
 - les variables sont stockées en lignes et en colonnes
- Données nettoyées (tidy data)
 - 1. chaque variable mesurée doit figurer dans une colonne
 - 2. Chaque observation différente d'une variable doit figurer sur une ligne différente
- Le passage raw data ->tidy data entraine une altération des données. Il faut toujours fournir
 - Les données brutes (pour comprendre)
 - Les données nettoyées (pour gagner du temps)
 - La description du processus de nettoyage (quels compromis)

Application

Naissances dans une maternité anglaise

- Données provenant de 500 naissances uniques (singleton births) dans un hôpital de Londres.
- Métadonnées
- Un tableau (data frame) de 500 observations concernant les 8 variables suivantes:

Variable	Signification
id	Identifiant pour la mère et l'enfant
bweight	Poids de naissance
lowbw	Poids inférieur à 2500 g (0 = non, 1 = oui)
gestwks	Durée de la gestation (semaines)
preterm	Durée de gestation inférieure à 37 semaines (0 = non, 1 = oui)
matage	Age maternel
hyp	Hypertension gravidique (0 = non, 1 = oui)
sex	Sexe de l'enfant (1 = garçon, 2 = fille)

Principaux paramètres statistiques

- Statistiques descriptives
 - Préalable à tout travail
 - Moyenne, variance, médiane, étendue
 - Généralement 1 variable à la fois (univarié)
- Aspect graphique des données
 - Paramétrique ou non
- Deux variables (bivarié)
 - Étude du croisement des variables

tests

	quantitatif	qualitatif
quantitatif	Coef. Corrélation Régression linéaire	
qualitatif	Boxplot ANOVA	Chi 2

- A partir de 3 groupes :
 - Analyse de la variance
 - Ex comparaison de l'age des femmes présentant une hémorragie au 1^{er}, 2ème et 3ème trimestre

Births

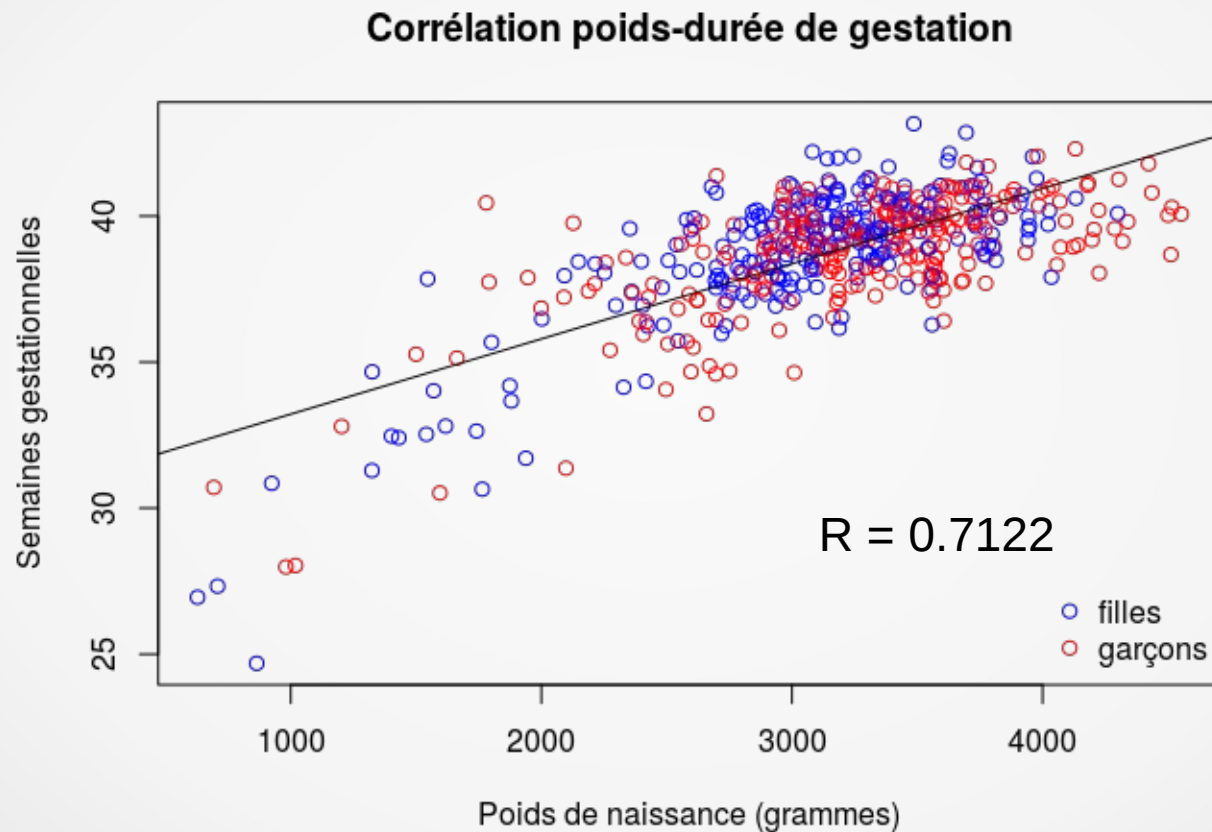
Data

	id	bweight	lowbw	gestwks	preterm	matage	hyp	sex
1	1	2974	0	38.52	0	34	0	2
2	2	3270	0	NA	NA	30	0	1
3	3	2620	0	38.15	0	35	0	2
4	4	3751	0	39.80	0	31	0	1
5	5	3200	0	38.89	0	33	1	1
6	6	3673	0	40.97	0	33	0	2

Statistiques univariées :

	id	bweight	lowbw	gestwks	preterm	matage	hyp	sex
Min.	: 1.0	Min. : 628	0:440	Min. :24.69	0 :427	Min. :23.00	0:428	1:264
1st Qu.	:125.8	1st Qu.:2862	1: 60	1st Qu.:37.94	1 : 63	1st Qu.:31.00	1: 72	2:236
Median	:250.5	Median :3188		Median :39.12	NA's: 10	Median :34.00		
Mean	:250.5	Mean :3137		Mean :38.72		Mean :34.03		
3rd Qu.	:375.2	3rd Qu.:3551		3rd Qu.:40.09		3rd Qu.:37.00		
Max.	:500.0	Max. :4553		Max. :43.16		Max. :43.00		
				NA's :10				

Exemple de corrélation



Ne pas confondre corrélation et relation de cause à effet
(QI et pointure)

Comparaison de 2 moyennes : principe des tests d'hypothèse

Q : le poids de naissance des garçons est-il plus élevé que celui des filles ?

On part d'une hypothèse neutre (hypothèse nulle) :

« il n'y a pas de différence de poids à la naissance entre les garçons et les filles »

Test de Student t-test

```
data:  births$bweight by births$sex
```

```
t = 3.4916, df = 492.914, p-value = 0.0002617
```

```
Hypothèse alternative: true difference in means is greater than 0
```

```
Intervalle de confiance à 95 %:
```

```
86.17495 307.96706
```

```
Poids moyen Garçons  
3229.902
```

```
Poids moyen fille  
3032.831
```

Conclusion : p est inférieur à 5 % (0,05). On rejette l'hypothèse nulle et on retient l'hypothèse alternative : en moyenne, le poids de naissance des garçons est plus élevé que celui des filles.

Deux variables qualitatives

Q : un poids faible à la naissance est indépendant de présence d'une HTA chez la mère ?

Poids faible	HTA gravidique		Sum
	non	oui	
non	388	52	440
oui	40	20	60
Sum	428	72	500

Pearson's **Chi-squared** test with Yates' continuity correction
data: t
X-squared = 18.12, df = 1, **p-value = 2.073e-05**

Conclusion : les 2 variables ne sont pas indépendantes. L'existence d'une HTA chez la mère a une influence sur le poids de l'enfant.

2 var. qualitatives dichotomiques

		HTA gravidique		
Poids	faible	non	oui	Sum
non		388	52	440
oui		40	20	60
Sum		428	72	500

		HTA gravidique		
Poids	faible	non	oui	Sum
non	VN	FP	440	
oui	FN	VP	60	
Sum	428	72	500	

- **Sensibilité** = $VP/(VP+FN) = 0,33$
 - Probabilité d'avoir un test positif si la maladie est présente
- **Spécificité** = $VN/(VN+FP) = 0,77$
 - Probabilité d'avoir un test négatif si la maladie est absente
- **VPP** = $VP/(VP+FP) = 0,28$
 - probabilité que la maladie soit présente lorsque le test est positif
- **VPN** = $VN/(VN+FN) = 0,79$
 - probabilité que la maladie ne soit pas présente lorsque le test est négatif

Odds Ratio (var.qualitatives)

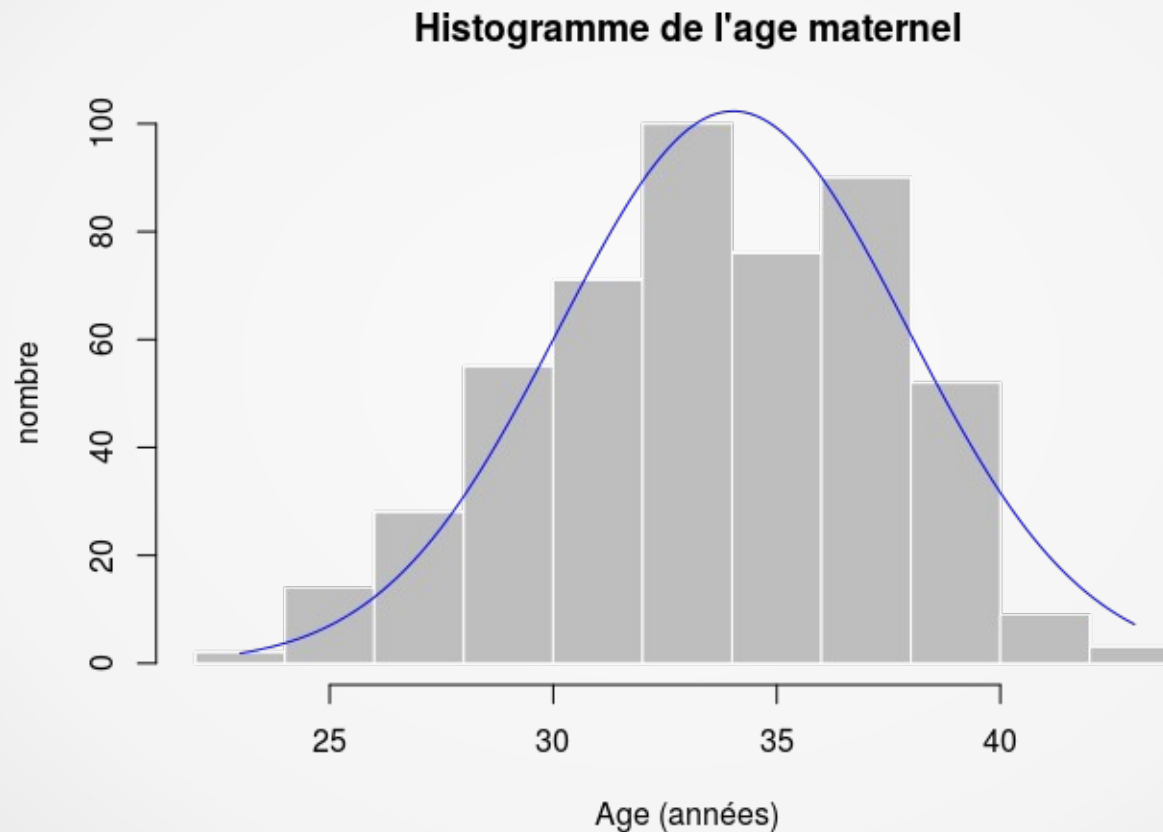
		cancer		
		oui	non	
fumeur	oui	1350	1296	2646
	non	7	61	68
		1357	1357	2714

- $p(K/f) = 1350/2646 = 0,51$
- $p(K/nf) = 7/68 = 0,10$
- $Od1 = 0,51/(1-0,51) = 1,04$
- $Od2 = 0,10/(1-0,10) = 0,11$
- $OR = 1,04/0,11 = 9,45$
- La probabilité de développer un cancer du poumon est 9,5 fois supérieure chez les fumeurs.

Notion de hasard

- Jet d'une pièce : la même pièce donne des résultats différents
- Lois statistiques
 - Loi normale (Gauss - Laplace)
 - Loi de Student (Gosset)
 - Loi de Poisson

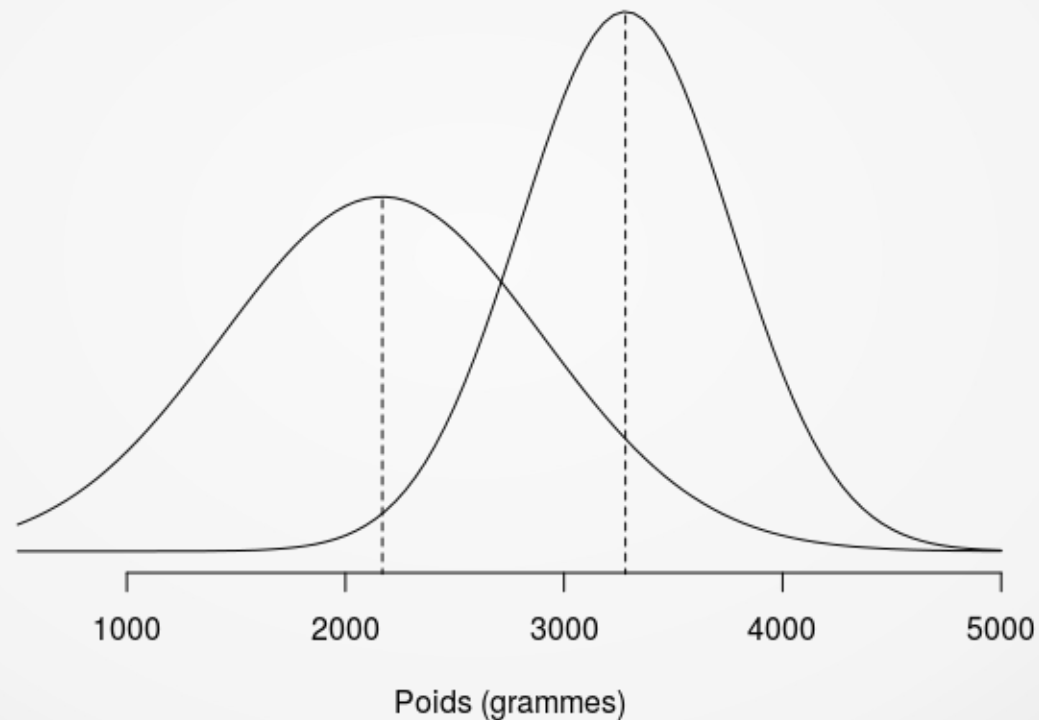
La loi normale



Deux paramètres : moyenne et variance permettent de résumer des centaines d'observations.

Sensibilité et spécificité

Poids des enfants prématurés et nés à terme



En fonction du poids peut-on dire si un enfant est né à terme ou pas ?

Population et échantillon

- Etudes prospectives rétrospectives, cas témoins, cohorte
- Simple et double aveugle



Merci pour votre attention

SAV : jeanclaude.bartier@gmail.com