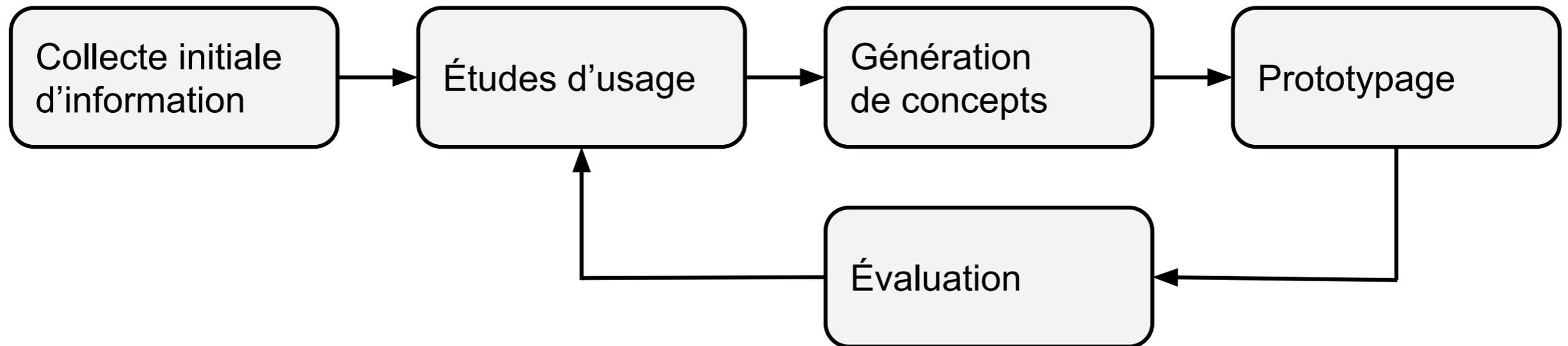


INF03 Expérience Utilisateur

6. Évaluation et tests – 2

Aurélien Tabard

Rappels des dernières séances



Évaluation et tests

- ▶ Introduction
- ▶ Approches d'évaluation
- ▶ Méthodes analytiques
- ▶ Méthodes empiriques
- ▶ Évaluation 2.0 : passer à l'échelle

- ▶ **Design expérimental**

Design expérimental

- ▶ Introduction et exemples
- ▶ Les éléments de base d'une expérience
- ▶ Définition d'une expérience
- ▶ Conduite de l'expérience
- ▶ Récupération et nettoyage des données
- ▶ Analyse des données
 - ▶ Exploratoire
 - ▶ Statistique

Design expérimental

- ▶ **Introduction et exemples**
- ▶ Les éléments de base d'une expérience
- ▶ Définition d'une expérience
- ▶ Conduite de l'expérience
- ▶ Récupération et nettoyage des données
- ▶ Analyse des données
 - ▶ Exploratoire
 - ▶ Statistique

Expériences contrôlées

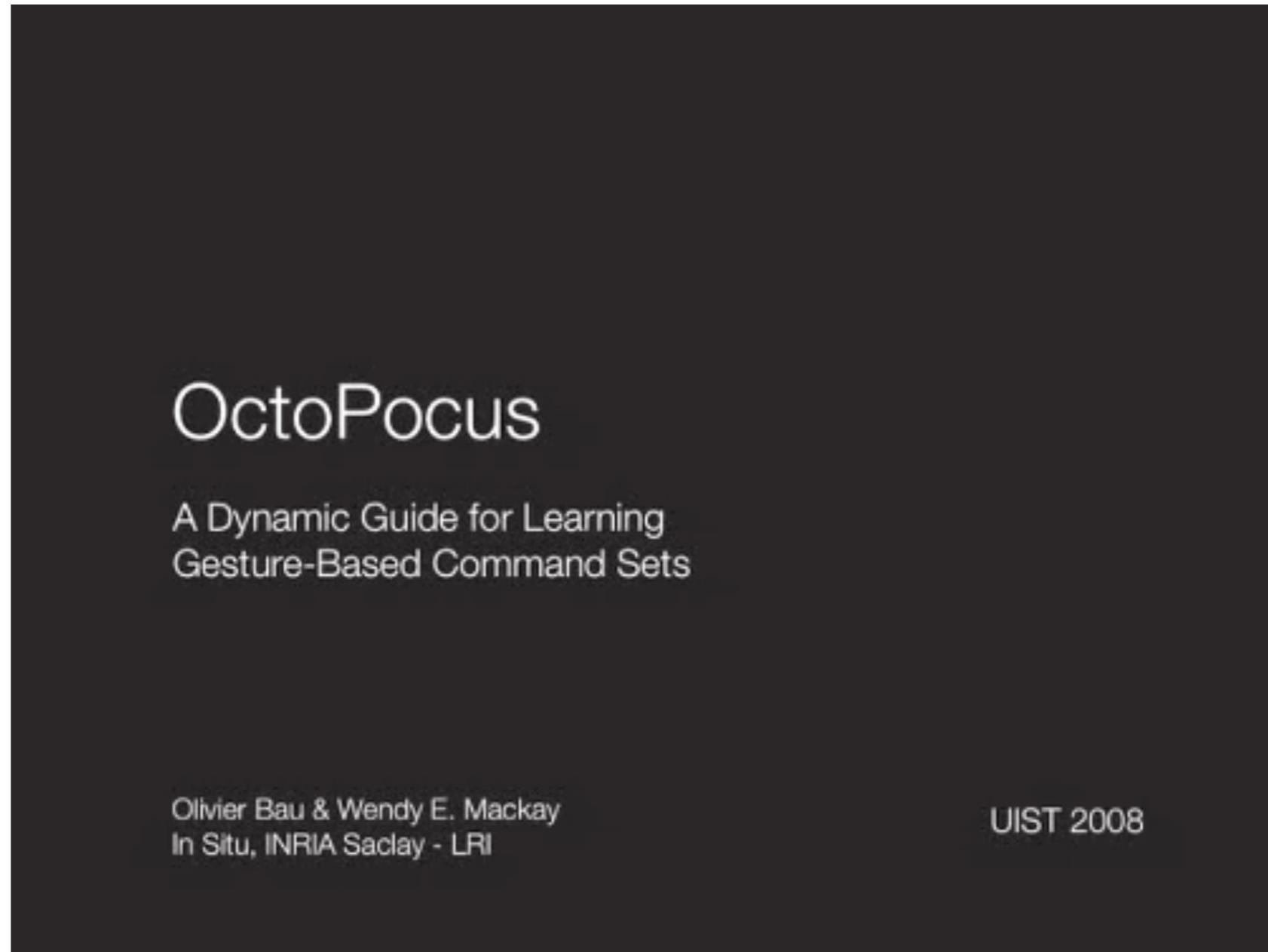
Une approche scientifique

- ▶ Réponses spécifiques à des questions
 - ▶ Performance
 - ▶ Apprentissage
 - ▶ Satisfaction
- ▶ Connaissance généralisables dans plusieurs contextes
- ▶ Montrer un lien de causalité
 - ▶ corrélation : montrer qu'un changement dans A entraîne un changement dans B
 - ▶ ordre : montre que A arrive avant B
 - ▶ pas de cause cachée : montrer qu'il n'y a pas de C avec $C \rightarrow A$ et $C \rightarrow B$

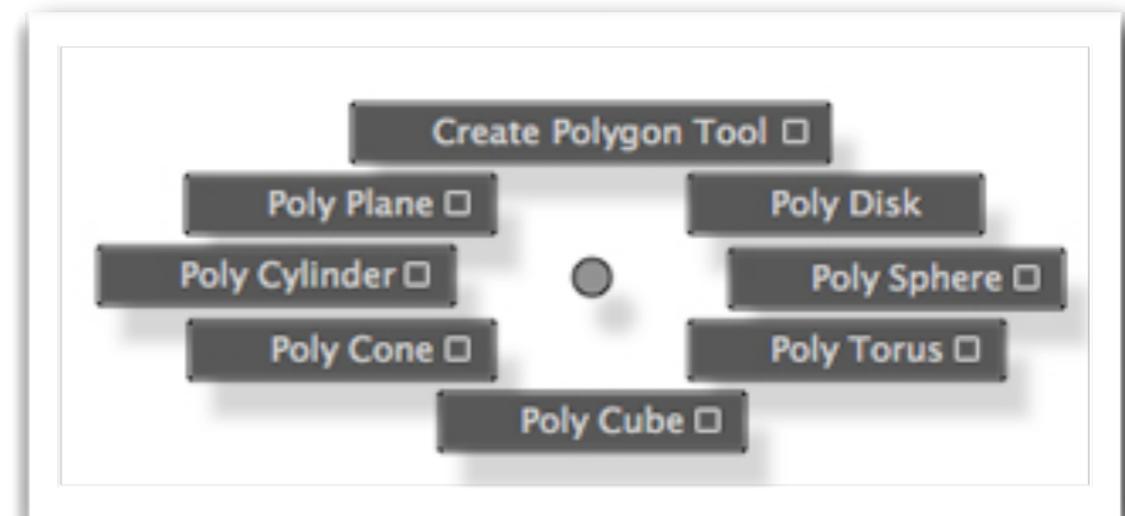
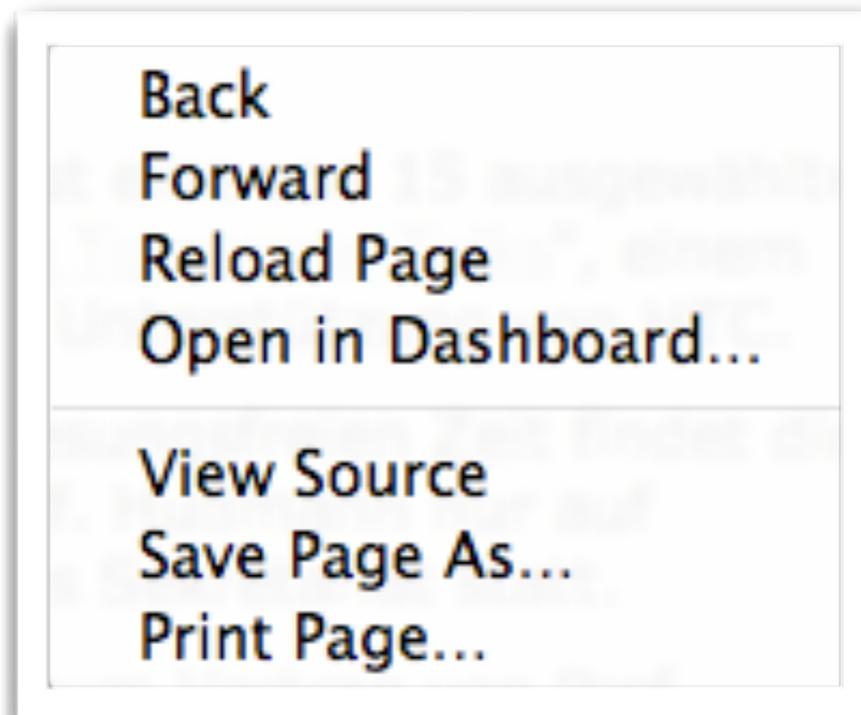
Exemple : Dispositifs de contrôle en entrée

Dispositif	Étude	IP (bits/s)
Main	Fitts (1954)	10,6
Souris	Card, English, & Burr (1978)	10,4
Joystick	Card, English, & Burr (1978)	5
Trackball	Epps (1986)	2,9
Touchpad	Epps (1986)	1,6
Eyetracker	Ware & Mikaelian (1987)	13,7

Exemple : Apprentissage de gestes



Exemple : comparer deux designs de menu



Design expérimental

- ▶ Introduction et exemples
- ▶ **Les éléments de base d'une expérience**
- ▶ Définition d'une expérience
- ▶ Conduite de l'expérience
- ▶ Récupération et nettoyage des données
- ▶ Analyse des données
 - ▶ Exploratoire
 - ▶ Statistique

Méthode 1.

1. Définir ce qu'on cherche : formuler une hypothèse
 - ▶ Le menu circulaire réduit le temps de recherche des éléments
2. Concevoir l'expérience, choisir variables et paramètres fixes
 - ▶ Définir la structure des menus
3. Conduire une expérience pilote pour tester l'expérience
 - ▶ Améliorer le design des menus

Méthode 2.

4. Sélectionner des participants

- ▶ Des étudiants passant plus de 2h / j sur leur ordinateur

5. Conduire l'expérience et collecter les données

6. Analyser les données pour accepter ou rejeter l'hypothèse et mesurer l'effet de la variation

- ▶ Temps de recherche moyen : 2.26 (Circulaire), 2.64 (Classique)
- ▶ La différence est significative : $p < .05$

Les éléments d'une expérience

- ▶ Facteurs (ou **variables indépendantes**)
 - ▶ Les variables qu'on va changer pour chaque condition
 - ▶ La quantité d'éléments dans un menu, ou le nombre de sous-menus
- ▶ Niveaux (les valeurs possibles des variables indépendantes)
 - ▶ Un menu avec 8 éléments ou un menu avec 12
- ▶ Mesures (ou réponses, ou **variables dépendantes**)
 - ▶ Les résultats mesurés de l'expérience.
 - ▶ Le temps de sélection d'un élément
- ▶ Réplication
 - ▶ le nombre de participants pour chaque niveau

Les variables indépendantes (facteurs)

- ▶ Les conditions d'une expérience sont définies par ces variables
 - ▶ The number of items in a list, text size, font, color
- ▶ Le nombre de variables différentes constitue le niveau
 - ▶ E.g., font can be times or arial (2 levels), background can be blue, green, or white (3 levels). This results in 6 experimental conditions (times on blue, times, on green, ..., arial on white)

Les variables dépendantes

Les variables dépendantes sont les variables qui vont être mesurées :

- ▶ Mesures objectives : e.g. temps pour réaliser une tâche, nombre d'erreurs, etc.
- ▶ Mesures subjectives : plaisir, frustration, etc.
- ▶ Elles ne devraient varier qu'en fonction de changements dans les variables indépendantes (= être fixes sinon).

Exercice

Identifier les variables indépendantes (facteurs) et les variables dépendantes (mesures) dans chaque scénario. Donner des niveaux possibles des variables indépendantes.

- ▶ Une étude pour voir si les gens ayant suivi une formation en sécurité utilisent des mots de passe plus sécurisés.
- ▶ Une étude qui cherche à voir qui d'un joystick ou d'une souris est plus efficace pour sélectionner des cibles statiques ou des cibles qui bougent
- ▶ Une étude pour savoir si les équipes qui utilisent du video Hangout/Skype sont plus productives que celles qui utilisent seulement des chats textes.

Les participants / sujets

Intra-sujets (Within-subjects design) :

- ▶ Participants exposés à toutes les conditions
- + Besoin de peu d'utilisateurs (10 - 20)
- Effet d'apprentissage

Inter-sujets (Between-subjects design) :

- ▶ Les participants sont séparés en groupes et chaque groupe exposé à une condition (contrôle et traitement)
- + Pas d'effet d'apprentissage
- demande plus d'utilisateurs

Gérer les effets d'ordre

L'ordre de présentation des traitements peut avoir un effet sur les mesures :

- ▶ apprentissage
- ▶ fatigue
- ▶ contraste (le premier traitement se reporte sur les réponses du 2e traitement)

Solution

- ▶ repos entre les traitements
- ▶ Équilibrer (counterbalancing), mais ça peut devenir compliqué
- ▶ Carré Latin

Carré Latin

via <http://hci.rwth-aachen.de/~chat/StatLecture/prerequisite.pdf>

Chaque condition apparait dans chaque position possible

Chaque condition précède chaque autre condition une fois

Exemple pour 6 traitements :

1	A	B	F	C	E	D
2	B	C	A	D	F	E
3	C	D	B	E	A	F
4	D	E	C	F	B	A
5	E	F	D	A	C	B
6	F	A	E	B	D	C

Exercice

Quel type de design expérimental ?

- ▶ Une étude pour voir si les gens ayant suivi une formation en sécurité utilisent des mots de passe plus sécurisés.
- ▶ Une étude qui cherche à voir qui d'un joystick ou d'une souris est plus efficace pour sélectionner des cibles statiques ou des cibles qui bougent
- ▶ Une étude pour savoir si les équipes qui utilisent du video Hangout/Skype sont plus productives que celles qui utilisent seulement des chats textes.

Hypothèses

- ▶ Prédire le résultat d'une expérience
- ▶ Dire comment un changement dans les variables indépendantes va impacter les variables dépendantes

Approche classique

- ▶ Formuler une hypothèse de travail H_1
- ▶ Formuler une hypothèse nulle H_0
 - ▶ intuition (naive) : si H_0 est faux alors H_1 doit être vrai
- ▶ Conduire une expérience et l'analyse statistique qui va avec pour falsifier cette hypothèse
- ▶ Si l'analyse statistique montre une différence significative il est probable que le résultat de soit pas dû au hasard

Exercice

Quel H_0 et H_1 ?

- ▶ Une étude pour voir si les gens ayant suivi une formation en sécurité utilisent des mots de passe plus sécurisés.
- ▶ Une étude qui cherche à voir qui d'un joystick ou d'une souris est plus efficace pour sélectionner des cibles statiques ou des cibles qui bougent
- ▶ Une étude pour savoir si les équipes qui utilisent du video Hangout/Skype sont plus productives que celles qui utilisent seulement des chats textes.

Validité

Validité interne

- ▶ Validité des résultats en cas de réplication

Validité externe

- ▶ Confiance de généralisation des résultats au monde réel

Cas pratique

Comparaison de la vitesse d'écriture entre un clavier d'ordinateur et un clavier mobile.

- ▶ Quelles variables indépendantes (facteurs) ?
- ▶ Quelles variables dépendantes (mesures) ?
- ▶ inter- ou intra- participants ?
- ▶ Quelles hypothèses ?

Design expérimental

- ▶ Introduction et exemples
- ▶ Les éléments de base d'une expérience
- ▶ Définition d'une expérience
- ▶ **Conduite de l'expérience**
- ▶ Récupération et nettoyage des données
- ▶ Analyse des données
 - ▶ Exploratoire
 - ▶ Statistique

Récupération des données

Observations de ce qu'on fait les utilisateurs

Données de logs

- ▶ Données des variables dépendantes (temps, erreurs)

Structure en tableau :

userid	group	condition	executiontime	error

Cas pratique

- ▶ Quel protocole ?
- ▶ Quel format des données collectées

Allez voir

- ▶ <http://10fastfingers.com/> pour vous donner une idée.

Design expérimental

- ▶ Introduction et exemples
- ▶ Les éléments de base d'une expérience
- ▶ Définition d'une expérience
- ▶ Conduite de l'expérience
- ▶ **Récupération et nettoyage des données**
- ▶ Analyse des données
 - ▶ Exploratoire
 - ▶ Statistique

Récupération des données

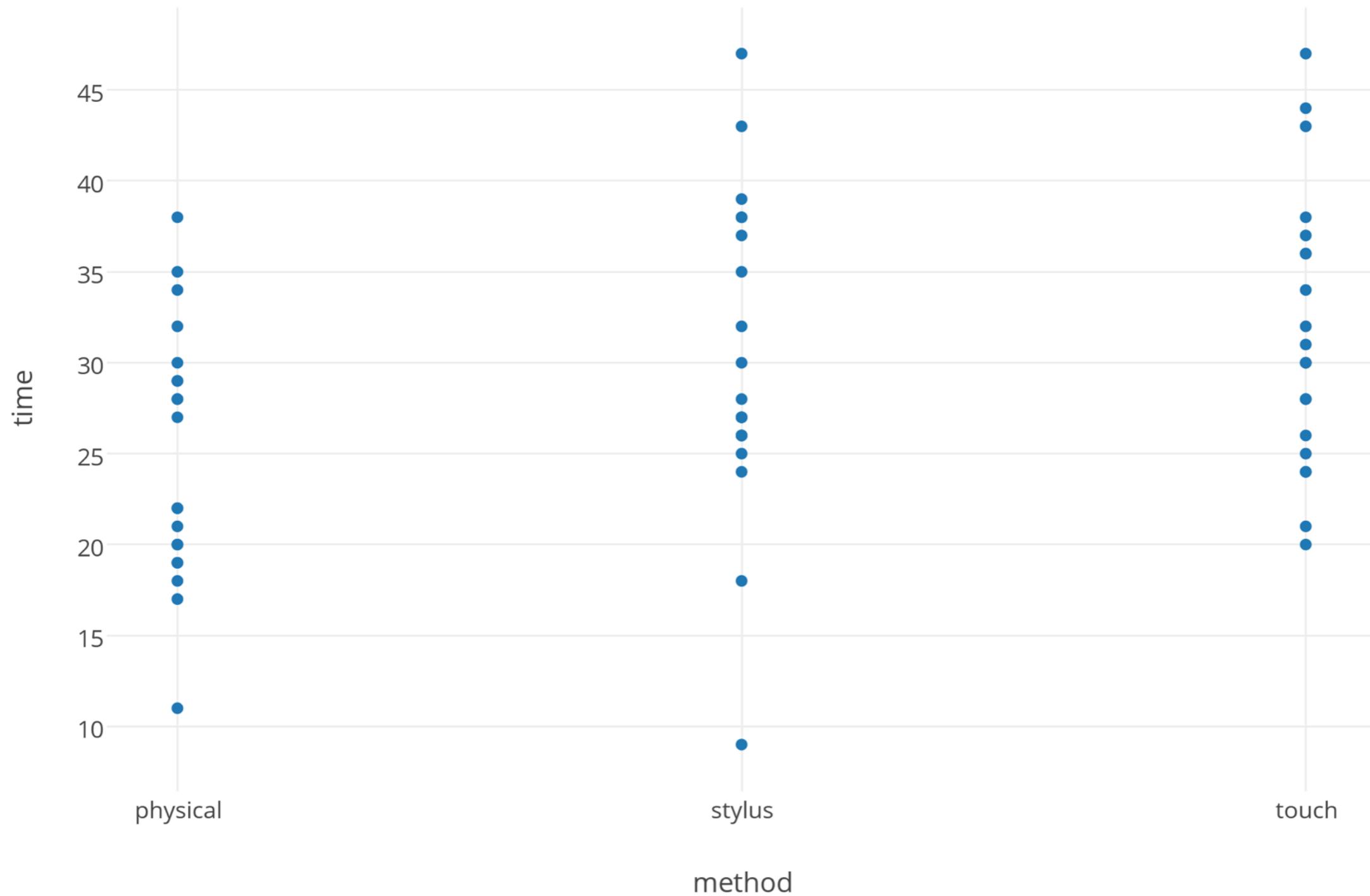
Agréger les données de tous les groupes dans un pad avec un format csv (comma separated value).

Les copier sur votre comptes plot.ly

Design expérimental

- ▶ Introduction et exemples
- ▶ Les éléments de base d'une expérience
- ▶ Définition d'une expérience
- ▶ Conduite de l'expérience
- ▶ Récupération et nettoyage des données
- ▶ **Analyse des données**
 - ▶ Exploratoire
 - ▶ Statistique

Qui est le plus rapide ?



Qui est le plus rapide

Ça dépend de savoir

- ▶ la différence entre les médianes
- ▶ la distribution des données (la déviation standard)
- ▶ la taille de l'échantillon
- ▶ si les moyennes sont significativement différentes

> Premier regard sur les données avec quelques graphes et des statistiques de base pour se faire une idée.

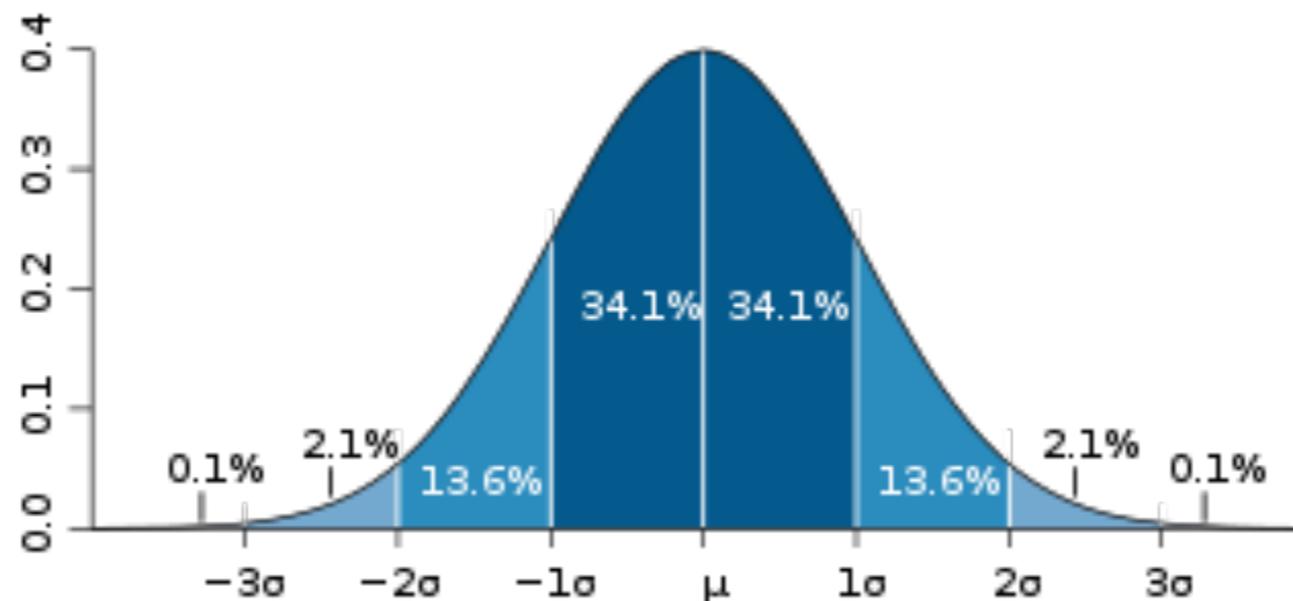
- ▶ partie exploratoire

(Student's) t-test

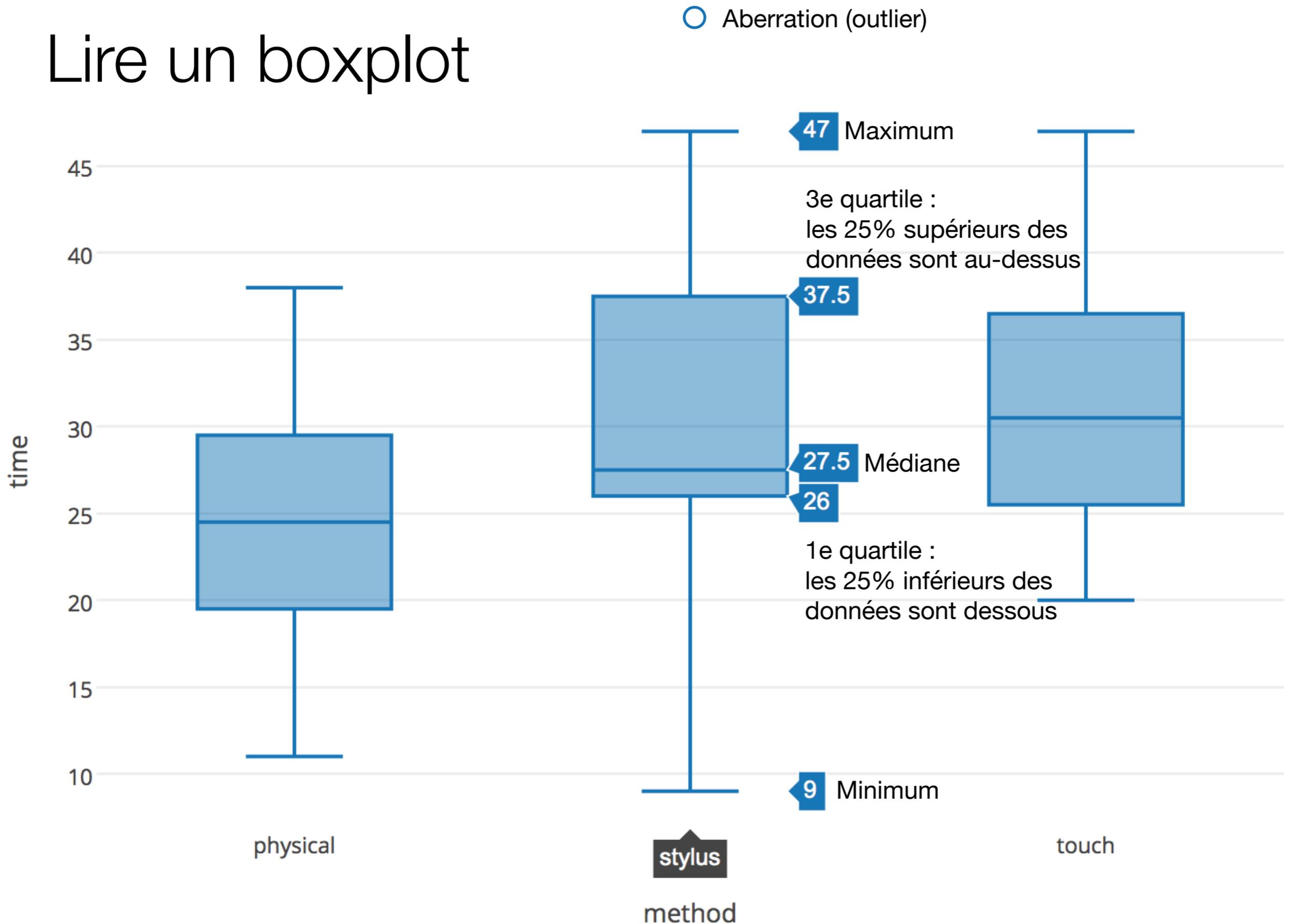
On regarde la relation entre deux jeux de données

Conçu pour :

- ▶ un petit échantillon (= peu de mesures)
- ▶ une déviation standard (et une moyenne) indéfinie
- ▶ mais une **distribution normale**



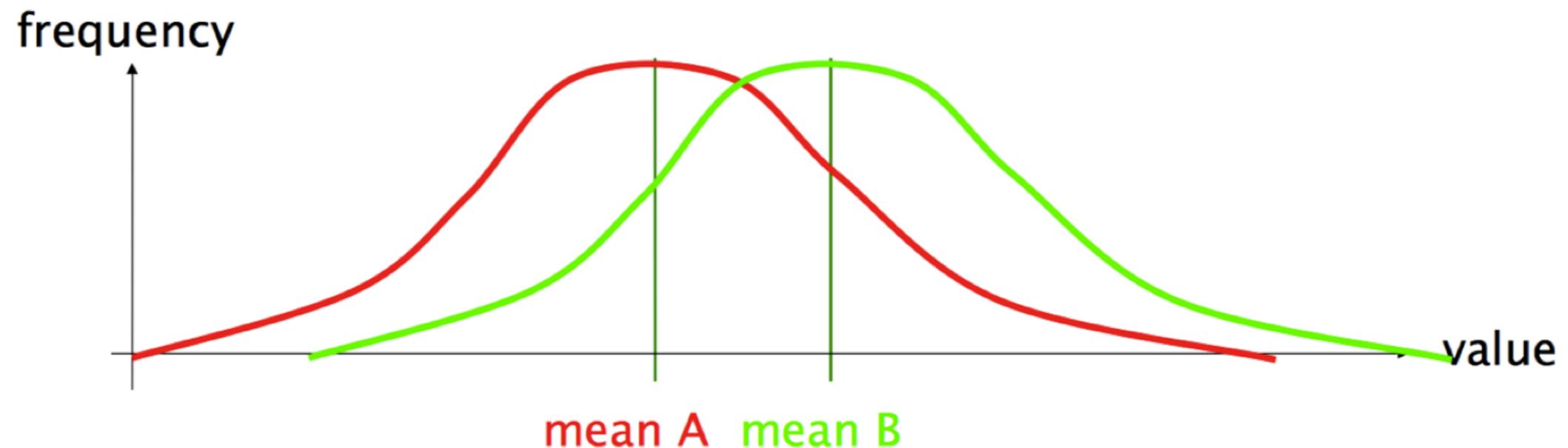
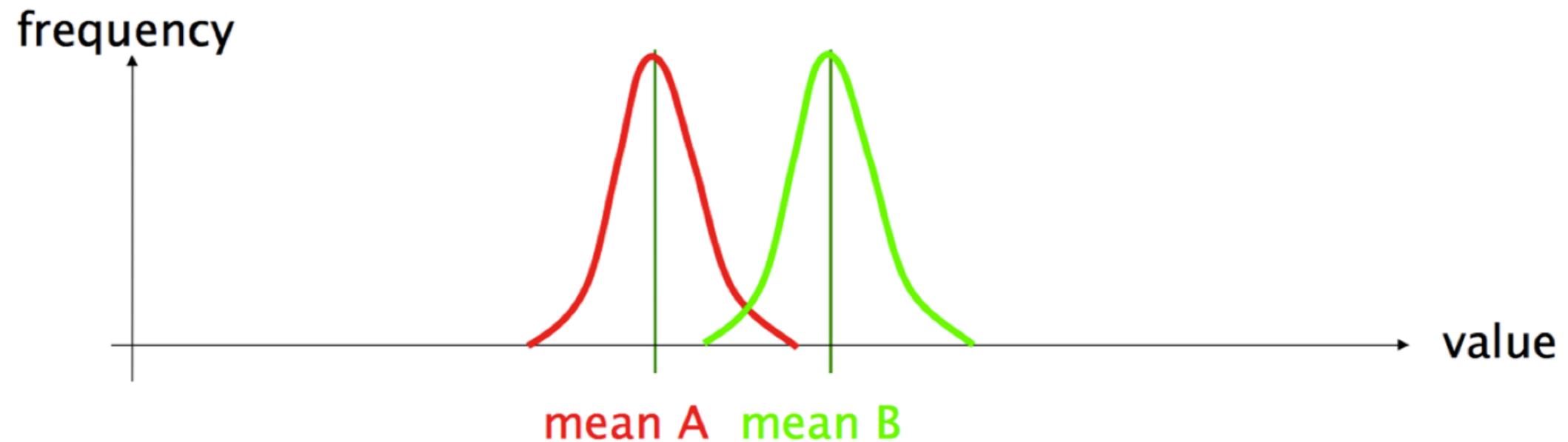
Lire un boxplot



Comparer des valeurs

via <http://www.medien.ifi.lmu.de/lehre/ws1213/mmi2/uebung/slides10.pdf>

Y a t'il une différence significative entre deux mesures ?



t-test

Donne p :

- ▶ **la probabilité que les deux populations aient la même moyenne**
- ▶ Et non la probabilité que le résultat soit par hasard...

En UX :

- ▶ $p < 0.05$ (= 5% de probabilité) est la convention (ou 0.01)
- ▶ un p plus faible (ex : 0.00001) ne veut pas dire que le résultat est “plus” significatif.
- ▶ un résultat significatif est différent d'un résultat est important

NE PAS

Si $p > 0.05$ dire :

- ▶ *“notre test a montré qu’il n’y avait pas de différence”*.
- ▶ différence significative -> il s’est passé quelque chose
- ▶ pas de différence significative -> rien

On ne peut pas montrer qu’il n’y a pas de différence
(avec les outils statistiques dont j’ai parlé)

t-statistique