

Cours 10: évaluation partie 2

un peu avancée, parties pas
pertinents pour vos examen
sont grisés

Anastasia.Bezerianos@lri.fr

(plusieurs slides sont basés sur des slides de N. Roussel, O. Chapius, T. Tsandilas, W. Mackay, M. Beaudouin Lafon, et S. Greenberg)

Evaluation d'un système

Techniques informels et rapides :

Heuristiques

Evaluation Heuristique

Design Walkthrough

Techniques formels :

Etude des alternatives aux utilisateurs

Expériences contrôlées

Quasi-expériences

Autres ...



évaluation : formelle et ciblée

Étude des alternatives aux utilisateurs

Tester plusieurs alternatives du système avec les utilisateurs :

techniques d'interaction

layouts d'écrans

systèmes d'aide

alternatives de conception

menus pop-up vs déroulant

hypertexte vs hiérarchie

tutor vs FAQs vs recherche

...

Etude des alternatives aux utilisateurs

(parfois connu comme « étude d'usabilité », à ne pas confondre avec l'évaluation heuristique qui évalue que l'usabilité)

But : Déterminer le meilleur choix de conception, avec les utilisateurs

Procédure

Décrire l'objectif de conception et plusieurs méthodes d'implémentation (ex., présentation ou interaction)

Choisir les variables indépendantes & dépendantes

indépendants : alternatives/choix, dépendantes : critères (ex. temps)

Faire une prédiction & spécifier l'hypothèse nulle (plus tard)

préd.: choix1 plus vite, hyp. nulle: tous les choix prennent le même temps

Préparer l'environnement pour tester chaque condition/choix

Utiliser au moins 3 sujets

Analyser les résultats

Les différences sont-elles significatives ? (plus tard)

Etude des alternatives aux utilisateurs

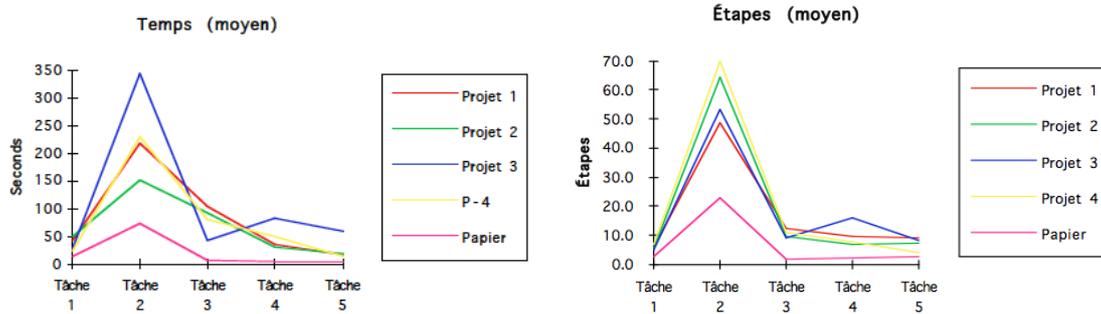
Toujours avec les vrais utilisateurs ...

Problèmes

- Comment définir une étape (quoi demander aux utilisateurs à faire) ?
- Et si l'utilisateur ne peut pas faire un tâche ?
- Que faire si le système est mal installé ?
- Quelle est l'importance des connaissances préalables de l'utilisateur ?
- Est-ce que ça marche différemment dans les contextes différents?

Etude des alternatives aux utilisateurs

Exemple de résultats (variables dépendants):
temps, étapes, erreurs, etc



Expériences contrôlées

But : Est-ce que le traitement X cause l'effet Y?
(plus ciblées que les études, facteurs testés plus contrôlés)

Facteur clé : Affectation aléatoire à des groupes

Test d'hypothèse

Contrôler les conditions pour isoler les variables

Comparer des hypothèses alternatives

Analyse de corrélations (connexions entre variables)

Mesurer le degré de corrélation entre deux facteurs

Connaître l'un aide à prédire l'autre

Expériences contrôlées : exemple

Exemple d'hypothèse:

Comparer les menus linéaires et circulaires, nous pensons que circulaires sont plus vites

HO (Hypothèse nulle) :

Il n'y a pas de différence de performance entre les utilisateurs en temps et taux d'erreur pour la sélection d'un item dans un menu linéaire ou un menu circulaire, quelle que soit l'expérience antérieure d'utilisation de la souris ou d'autres types de menus par l'utilisateur.

Expériences contrôlées : exemple

Spécifier les variables indépendantes

Les variables indépendantes (= facteurs) sont celles que l'on fait varier ou que l'on contrôle, indépendamment l'un de l'autre

Variables indépendantes :

- 2 Types de menu : linéaire, circulaire
- 5 Nombre d'items : 3, 6, 9, 12, 15
- 3 niveaux d'expertise: expert, novice, intermittent => $2 \times 5 \times 3 = 30$ conditions à tester

Expériences contrôlées : exemple

Spécifier les variables dépendantes

Les variables dépendantes (= **mesures**) sont celles que l'on mesure : ils dépendent du comportement du sujet

Pour faire une analyse statistique fiable, il faut suffisamment de mesures pour chaque condition

Variables dépendantes typiques en IHM :

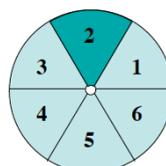
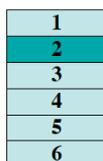
- Temps pour sélectionner un item
- Nombre d'erreurs
- Autres ?

Expériences contrôlées : exemple

Opérationnaliser le comportement
isoler ce que nous voulons tester

Dans notre expérience :

- Mêmes étiquettes pour les items des menus
- Même position du menu (centre de l'écran)
- Afficher l'item à sélectionner au lieu d'avoir à le trouver



Expériences contrôlées : exemple

Conduire l'expérience

Prédiction \neq Hypothèse Nulle

Hypothèse Nulle : inverse de notre prédiction qu'on peut réfuter (avec des méthodes statistiques)

Pour notre expérience :

- La performance des les deux types de menu est la même quelle que soit l'expertise, la taille du menu, et le rang de l'item

Expériences contrôlées : exemple

Conduire l'expérience

Faire signer un consentement éclairé (« **informed consent** ») aux sujets ← important en generale

Identifier les sujets et assurer leur anonymat

- Associer un numéro à chaque sujet
- En fonction de ce numéro, le programme doit connaitre les conditions pour ce sujet

Collecter les données expérimentales

- S'assurer qu'elles sont fiables et valides
- Minimiser les traitements lors de la collecte
- collecter des données brutes

Expériences contrôlées : problèmes

Problèmes de fiabilité

- Les résultats seraient-ils les mêmes si l'expérience était répétée ailleurs ?
- Qu'en est-il des différences individuelles ?
- Utiliser un nombre "raisonnable" de sujets (12 minimum, 30 approche une distribution normale)
- Utiliser des statistiques appropriées

Problèmes de validité

- L'expérience mesure-t-elle quelque chose de pertinent en dehors du laboratoire ?
- Les utilisateurs sont-ils typiques ?
- Les tâches sont-elles typiques ?
- L'environnement physique est-il différent?

Quasi-Expérience

Problème:

Comment faire de l'inférence causale dans un contexte réel ?

Situation:

Avoir des traitements, des mesures de résultats et des unités expérimentales
Affectation aléatoire impossible

Solution:

Rendre explicites les "menaces d'invalidité"

Quasi-Expérience

Identifier les sources potentielles d'interprétation erronée (ex. l'histoire, la maturation, les tests, la sélection des sujets, la perte des sujets, etc)

Triangler :
(voir le problème de différents angles)

Utiliser des méthodes d'analyse multiples, avoir des analyses des données en temps, avoir plusieurs personnes analyser les données

Autres

Techniques utilisés pour identifier les utilisateurs :

- Observation
- Interviews
- Questionnaires
- Focus group

Mais maintenant axé sur les aspects du nouveau système

évaluation :

un peu de statistiques

Analyser les données

Deux faces du même problème :
comment déterminer la réalité ?

Statistiques :
Sachant ce qu'il y a dans votre main, qu'y a-t-il dans le seau ?

Probabilité :
Sachant ce qu'il y a dans le seau, qu'y a-t-il dans votre main ?

Analyse statistique

Fournit les caractéristiques mathématiques des données

Décrit comment des ensembles de données se rapportent les uns aux autres

Estime la probabilité que les hypothèses soient correctes

Mesures standards en IHM

Temps pour exécuter une opération

Fréquences de certaines actions

Nombre d'erreurs

Reactions qualitatives
ex. "être perdu"

Statistiques

C'est un domaine TRÈS vaste

Il est difficile de faire des suppositions correctes

Il est facile de se tromper

Statistiques descriptives

Réduire la masse des données,

ex : la moyenne, la distribution

Statistiques inférentielles

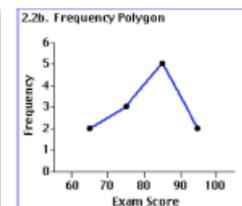
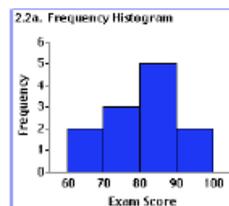
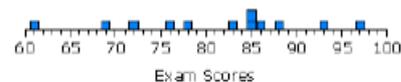
Inférer des propriétés d'une population à partir de celles d'un échantillon,

ex : mesurer la probabilité qu'une différence observée soit réelle

Statistiques descriptives

Comment résumer un ensemble de mesures d'une variable

Distribution des fréquences des valeurs



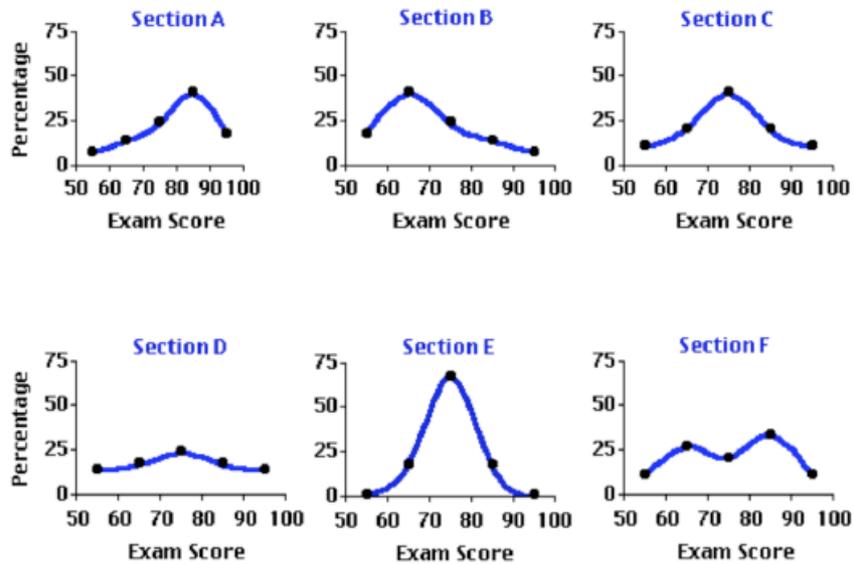
Types de distributions

Mesures de tendance centrale

Mesures de variabilité

Mesure de la corrélation entre deux variable

Types de distribution



Mesures de tendance centrale

S'appliquent à des variables scalaires (numeros)

Moyenne : Somme des valeurs divisée par leur nombre

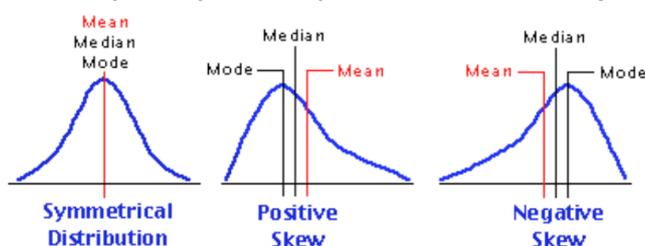
Médiane : Valeur "milieu" des N valeurs triées

N impair : valeur d'indice $(N+1)/2$

N pair : moyenne des valeurs d'indice $N/2$ et $N/2 + 1$

Mode : Valeur la plus fréquente

Il peut y avoir plusieurs modes (2 modes=bimodal)



Measures of variability

Mesurent l' "étalement" de la distribution

Intervalle : distance entre valeur minimale et maximale

Variance et écart-type :

somme des carrés entre chaque valeur et la moyenne

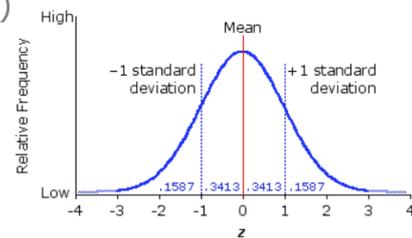
$$\text{variance} : s^2 = \frac{\sum (X_i - M)^2}{N}$$

$$\text{écart-type} : s = \sqrt{\frac{\sum (X_i - M)^2}{N}}$$

L'écart type peut servir

d'échelle de la distribution

$$z_i = \frac{(X_i - M)}{s}$$



Corrélation entre deux variables

Mesurer la relation entre deux variables scalaires

En général une variable indépendante X et une variable dépendante Y

Mesure du coefficient de corrélation linéaire r ($-1 \leq r \leq 1$)

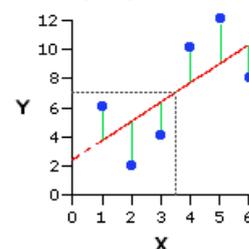
$$r = \frac{\sum (X_i - MX)(Y_i - MY)}{\sqrt{\sum (X_i - MX)^2 \sum (Y_i - MY)^2}}$$

r^2 s'interprète comme la proportion de la variabilité de Y qui est associée à la variabilité de X

$1 - r^2$ est la variance résiduelle, celle qui n'est pas expliquée

ATTENTION :

corrélation ne veut pas dire **cause**



Statistiques inférentielles

Complexe, plus puissant que les stat. descriptives
Fondé sur la théorie des probabilités

Ex: Comparing Means
Student test (t-test), ANOVA

Ex.: correlation
Pearson rho factor

Ex.: Regression Analysis

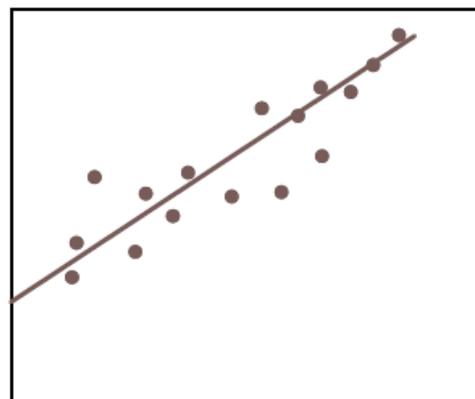
Regression

Calcule la droite qui approche au mieux les données

$y = .99x - 1.32$
 $r^2 = .668$

Utiliser la valeur d'une variable pour prédire l'autre

C
o
n
d
i
t
i
o
n
2



Condition 1

Signification statistique

P-value : Critère alpha = 0,05

0,05 = 1/20

S'il n'y a pas de différence et
que je fait cette expérience 20 fois,

un test donnera un résultat significatif
les 19 autres donnera un résultat non-significatif
(5% chance que l'observation est aleatoire)

Signification statistique

Fournit une estimation quantitative de la probabilité
que deux distributions soient différentes

Signification ≠ importance !!!!

Si le nombre de sujet est grand, une petite différence
peut produire un résultat significatif