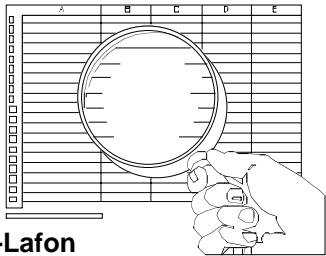


1

Analyse quantitative de données

- Statistiques
- Analyse Exploratoire de Données



Michel Beaudouin-Lafon
Université Paris-Sud

Cours de Wendy E. Mackay (I.N.R.I.A.)

2

Statistiques

- C'est un domaine TRÈS vaste
- Il est difficile de faire des suppositions correctes
- Les erreurs sont courantes
- On peut faire des tests simples

3

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

4

Mesures standards en IHM

- Temps pour exécuter une opération
- Fréquences de certaines actions
- Nombres d'erreurs
- Réactions qualitatives

par exemple : "être perdu"

1

Analyse de données

- Le type des données et le dispositif expérimental déterminent les statistiques applicables
- Echelles de données

Nominale	Maillots de footballers
Ordinale	Classements
Intervalle	Température
Ratio	Poids
- Statistiques descriptives ou inférentielles

2

Statistiques descriptives

- Plus simple mais moins puissant
- Réduit la masse des données
 - Intervalle des valeurs
 - Distribution de fréquences des valeurs
 - Mesures de tendances centrales
 - moyenne médiane mode
 - Mesures de variance autour du centre
 - variance déviation standard

3

Mesures de tendances centrales

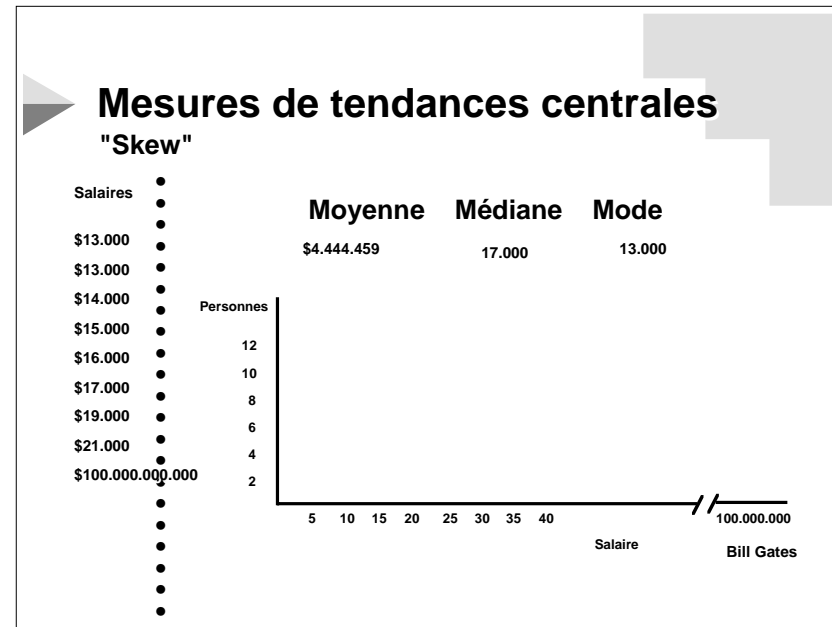
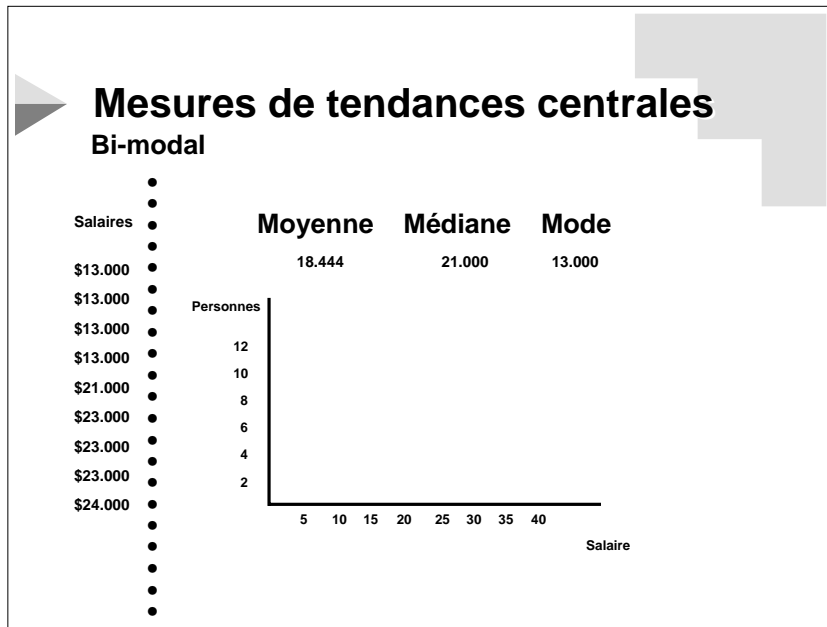
Normale, Gaussian, "Bell"

- Moyenne = Médiane = Mode
- Calculations < 50% > le plus souvent

4

Mesures de tendances centrales

Salaires	Moyenne	Médiane	Mode
\$12.000	14.000	14.000	14.000
\$13.000			
\$13.000			
\$14.000			
\$14.000			
\$15.000			
\$15.000			
\$16.000			



3

Statistiques inférentielles

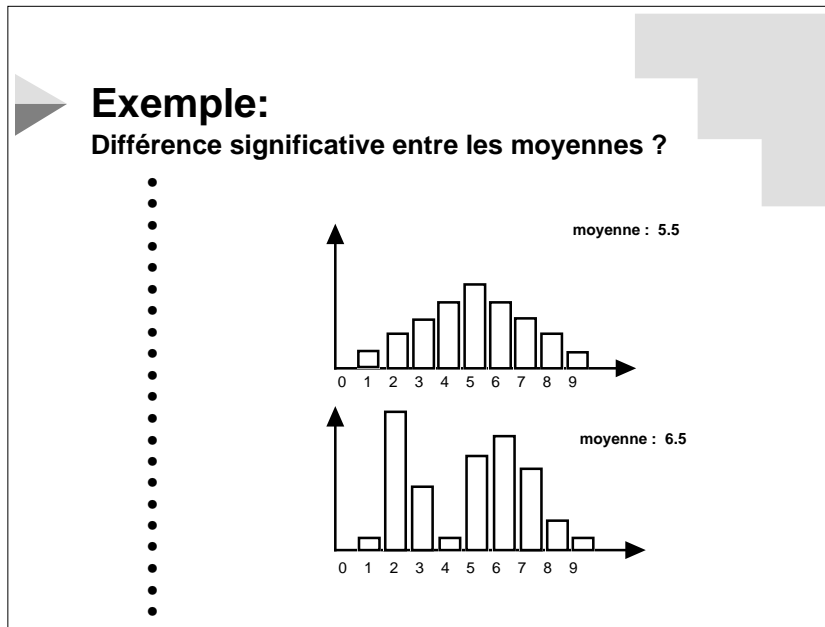
- **Complexe, plus puissant**
Fondé sur la théorie des probabilités
- **Exemple : Comparer les moyennes**
T test
- **Exemple : Mesurer une corrélation**
facteur rho de Pearson
- **Exemple : Analyse de régression**

4

Exemple :

Différences entre moyennes

- **Données :**
Deux ensembles de données qui mesurent le temps de sélection d'un item dans un menu
- **Question :**
Y a-t-il une différence réelle entre ces deux ensembles de données, ou bien les différences sont-elles seulement dues au hasard?
- **Hypothèse nulle :**
Il n'y a pas de différence entre les moyennes



1

2

Signification statistique

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

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

Inspection visuelle des données

- Les données sont toujours variables !
- Attention à ne pas tirer de conclusion erronée
- Distinguer une variation aléatoire d'une variation systématique
- **MAIS :**
- L'analyse exploratoire de données utilise des techniques visuelles pour comprendre les données

3

4

T-Test

- Le plus utilisé en IHM
- Donne le niveau de confiance que la différence entre deux moyennes est due à une différence réelle et non pas à une variation aléatoire
- **NE DIT RIEN** de la taille de la différence entre les moyennes

1

Exemples de T-Test :

- **Comparer 2 ensembles d'observations indépendantes**
 - Deux groupes avec des sujets différents
 - Le nombre de sujets par groupe peut varier
- **Observations par couples**
 - Un groupe avec des conditions différentes
 - Les couples de données d'un sujet sont comparés
- **Non-directionnel ou directionnel**
 - "Two-tailed test"
 - Différences dans les deux directions
 - "One-tailed test"
 - Différences dans une seule direction

2

Suppositions pour le T-Test

- **Les données sont distribuées aléatoirement**
 - Distribution normale
- **Les variances des populations sont égales**
 - Le test est cependant robuste si ce n'est pas le cas
- **Les observations individuelles sont indépendantes**
- **p-value: la probability que les moyennes ne soient pas différentes**
- **Choisir le critère au préalable ($p=.05$)**
- **Accepter or rejeter l'hypothèse nulle en fonction de la p-value**

3

Types d'erreur

Théorie de "Signal Detection"

		Réalité	
		Signal	Bruit
Prédiction	Signal	Hit	False Alarm
	Bruit	Miss	Correct Rejection

4

Niveau de signification et types d'erreur

- **Erreur de type 1**
 - Rejeter l'hypothèse nulle alors qu'elle est vraie
- **Erreur de type 2**
 - Accepter l'hypothèse nulle alors qu'elle est fausse
- **Un fort niveau de signification augmente les chances d'une erreur de type 2**
- **Un faible niveau de signification augmente les chances d'une erreur de type 1**

1

Corrélation

Condition 2

Condition 1

2

Corrélation

Mesure à quel point deux données sont corrélées

Fondée sur un coefficient de corrélation

- +1 Corrélation positive parfaite
- 0 Pas de corrélation
- 1 Corrélation négative parfaite

Corrélation ≠ Cause !!

Pas fiable avec des petits groupes

3

Régression

Calcule la droite qui approche au mieux les données

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

4

Exemple :

Condition	
1	2
5	6
4	4
6	6
4	4
5	5
3	3
5	5
4	4
5	5
6	6
6	6
7	7
6	6
7	7

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

Condition 2

Condition 1

1

Erreurs dans l'interprétation

- **TRES fréquentes : ce n'est pas facile !**
- **Signification \neq Importance**
- **Les suppositions sont souvent fausses :**
 - Les données ne suivent pas une distribution normale
 - Puissance du test
 - Valeur de la "p-value" \neq taille de l'effet
 - Signification vs. force de la relation

2

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**

3

Analyse exploratoire de données

Tukey & Mosteller

- **Processus interactif et itératif**
- **Explorer les données pour découvrir leurs propriétés :**
 - Le "doux" et le "rugueux"
- **Principes clés**
 - Etre ouvert
 - Etre sceptique

4

Analyse exploratoire de données

- **Importance de la visualisation**
 - Mais base mathématique solide
- **Mesures de :**
 - Position : moyenne, médiane, mode
 - Étalement : variance, déviation standard
 - Forme : courbe en cloche, biais, points isolés
- **Représentation graphique des données**
 - Graphes "Stem-and-leaf"
 - "Box-and-whisker"

1

Box-and-whisker plot

2

Études utilisateurs

Toujours avec les vrais utilisateurs !

Problèmes

- Comment définir une étape ?
- 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 ?

3

Intérêt des tests d'utilisabilité

Rapport coût-bénéfice

Number of user tests	Benefit-to-Cost Ratio
1	30
2	40
3	42
4	43
5	42
6	40
7	38
8	36
9	34
10	32
11	30
12	28
13	26
14	25
15	24
16	23
17	22
18	21
19	20
20	19

Nielsen & Landauer (1993)

4

Les problèmes d'utilisabilité

Combien de tests sont nécessaires ?

Number of tests	Easy to detect (d = .50)	Average to detect (d = .30)	Hard to detect (d = .15)
1	0.50	0.30	0.15
2	0.75	0.45	0.25
3	0.85	0.55	0.35
4	0.90	0.60	0.40
5	0.92	0.65	0.45
6	0.93	0.68	0.48
7	0.94	0.70	0.50
8	0.94	0.72	0.52
9	0.95	0.73	0.53
10	0.95	0.74	0.54
11	0.95	0.75	0.55
12	0.95	0.75	0.55
13	0.95	0.75	0.55
14	0.95	0.75	0.55
15	0.95	0.75	0.55
16	0.95	0.75	0.55
17	0.95	0.75	0.55
18	0.95	0.75	0.55
19	0.95	0.75	0.55
20	0.95	0.75	0.55

Nielsen & Landauer (1993)

d = détectabilité
 Probabilité qu'un test identifiera un problème

Résultats

Types d'observation

- Functionalité**
Pas d'aide, etc.
- Generale**
Difficile à utiliser, Trop d'icônes
- Specifique**
Il faut quitter pour effacer un postit
- Bugs**
L'import de postit ne fonctionne pas
- Suggestions**
Une vue generale sur tous les postits créés
- Décisions**
Tous les postits sont dans le meme mode

1

2

Résultats

Temps

Tâche	Projet 1	Projet 2	Projet 3	P-4	Papier
Tâche 1	~50	~50	~50	~50	~50
Tâche 2	~340	~220	~150	~100	~70
Tâche 3	~100	~100	~100	~100	~100
Tâche 4	~100	~100	~100	~100	~100
Tâche 5	~100	~100	~100	~100	~100

3

4

Résultats

Étapes

Tâche	Projet 1	Projet 2	Projet 3	Projet 4	Papier
Tâche 1	~10	~10	~10	~10	~10
Tâche 2	~68	~55	~45	~35	~25
Tâche 3	~10	~10	~10	~10	~10
Tâche 4	~10	~10	~10	~10	~10
Tâche 5	~10	~10	~10	~10	~10