Programmation des Interfaces Interactives Avancées

Ouriel Grynszpan

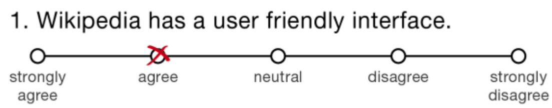
Professeur, Université Paris-Saclay

LISN lisn.upsaclay.fr

Évaluation - expérimentation

Collecte de données qualitatives

- Entretiens:
  - Non-structuré
  - Semi-structuré
  - Structuré (questions à choix fermé)
- Questionnaires
  - Échelle de Likert



- Observation en environnements contrôlés : technique du “think-aloud”
- Observation:
  - Directe (dans l’environnement naturel de l’utilisateur)
  - Indirecte (logging d’interaction)

Il existe différentes méthodes pour collecter des données qualitatives sur l’utilisation d’une IHM: les entretiens, les questionnaires ou l’observation. Cette dernière peut être en environnement contrôlé et suivre un protocole strict, prévu à l’avance. Elle peut aussi être effectuée directement dans l’environnement de l’utilisateur, pendant qu’il utilise l’IHM, ou de manière indirecte, en se basant sur des fichiers logs générés pendant l’interaction avec le système.

Analyse qualitative des données

- Catégorisation des données issues des réunions, entretiens ou protocoles “think-aloud”
- Schéma de codage
- Évaluation par plusieurs juges → fiabilité inter-juge
  - e.g. Kappa de Cohen: $\kappa = \frac{P(a) - P(e)}{1 - P(e)}$

P(a) : pourcentage d’accord observé
P(e) : Probabilité d’accord attendue due à la chance (e.g. 0.5)

Les données qualitatives peuvent la plupart du temps être transformées en données quantitatives. Par exemple, il est courant de transcrire les entretiens et témoignages des utilisateurs, puis de les analyser en suivant un schéma de codage. Le schéma de codage va associer un code à différents segments du texte transcrit, en fonction par exemple des thèmes abordés. Les codes peuvent ensuite être dénombrés aboutissant à des données quantitatives. Pour plus d’objectivité, l’analyse peut être réalisée par plusieurs juges. Il existe des métriques statistiques permettant d’évaluer l’accord entre les juges, ce qui permet de valider la pertinence du schéma de codage.

Données quantitatives

- Taux de succès
- Temps de réponse
- Fréquences d'un évènement (e.g. fréquence d'utilisation d'une fonctionnalité)
- Mesures physiologiques :
  - Temps de visualisation (eye-tracking)
  - Conductance cutanée
  - Fréquence cardiaque
  - Dilatation de la pupille
- Métriques issues de l'analyse qualitative

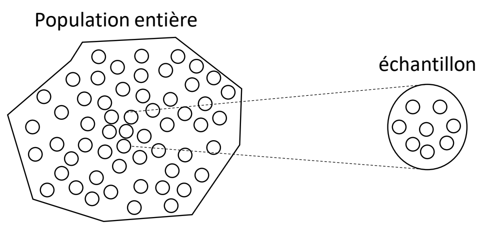
D'autres données qualitatives peuvent également être collectées directement pendant l'interaction avec le système. Voir la diapositive pour des exemples. Les données quantitatives peuvent ensuite être traitées statistiquement afin d'obtenir un résultat qui pourra être généralisé, c'est-à-dire valide pour toute utilisation similaire à celle qui aura été testée. C'est ce qu'on appelle les statistiques inférentielles.

Analyse statistique inférentielle

Lorsqu’on réalise un test avec des utilisateurs, il est difficile de savoir si les résultats peuvent être généralisés ou s’ils ne sont valables que pour le test effectué. Dans ce dernier cas, le test serait peu informatif. Nous utilisons des statistiques inférentielles pour savoir si les résultats peuvent être généralisés. La suite de ce cours vous propose une introduction à ces notions.

Objectif

- Inférer les propriétés générales d'une population à partir de celles observées sur un échantillon



- Représentatif
- Tirage aléatoire des individus de la population
- Variation aléatoire des données

Lorsqu'on fait un test utilisateur pour une application grand public, on ne teste qu'un échantillon limité d'utilisateurs potentiels. Pour que le test ne soit pas biaisé, il faut que cet échantillon soit représentatif, c'est à dire que les individus de différentes catégories soient dans des proportions similaires à celles de la population générale. Par exemple, les proportions de personnes jeunes ou âgées doivent correspondre à la population générale. Il faut en outre que l'échantillon ait été sélectionné au hasard. On ne doit pas favoriser certains individus plutôt que d'autres. Il est important de comprendre que chaque échantillon produira des données différentes. Donc on n'aura jamais exactement les mêmes résultats si on réalise le test plusieurs fois. Les données varieront de manière aléatoire.

Variable aléatoire

- Définition : quantité variable pouvant prendre un ensemble de valeurs discrètes ou continues sans que l'on puisse en connaître la valeur exacte avant observation.
- Exemples:
  - Durée d'un sprint
  - Position d'un électron
  - Temps de réponse à une question
  - Fréquence cardiaque …
- On va s'intéresser à la probabilité qu'une variable aléatoire tombe dans un intervalle de valeurs donné.

Nous devons donc considérer les données obtenues lors d'un test utilisateur comme des variables aléatoires. Nous n'allons donc pas chercher à déterminer des valeurs exactes, mais plutôt la probabilité que les données soient dans un intervalle de valeurs donné.

Intervalles intéressants

- Soit X une variable aléatoire de moyenne μ et écart type σ
- Soit $\bar{X}$ la moyenne de *n* mesures de cette variable calculée sur un échantillon
- Intervalle de variation: Intervalle du type $[\mu - l\ ,\ \mu + l]$, centré sur μ dans lequel $\bar{X}$ a 95% de chance de se trouver
- Intervalle de confiance: Intervalle du type $\left[\overline{X} - l\ ,\ \overline{X} + l\right]$, centré sur $\bar{X}$ dans lequel μ a 95% de chance de se trouver
- On va chercher à déterminer l'intervalle de confiance
- L'intervalle de confiance peut se déduire de l'intervalle de variation

En général, on ne connaît pas la vraie valeur de μ mais on peut calculer la moyenne de X sur un échantillon. L'intervalle de confiance donne un champ de valeurs probables pour μ.

Paramètres d’une variable aléatoire

- Moyenne ou Espérance :
  - Variable continue : $E(X) = \mu = \int_{-\infty}^{+\infty} xp(x)dx$
  - Variable discrète : $E(X) = \mu = \sum_{i=1}^{n} x_i p(x_i)$
- Variance :
  - Variable continue : $\mathrm{Var}(X) = E\left((X-\mu)^2\right) = \sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 p(x)dx$
  - Variable discrète : $\mathrm{Var}(X) = E\left((X-\mu)^2\right) = \sigma^2 = \sum_{i=1}^{n} (x_i-\mu)^2 p(x_i)$
- Écart Type: $\sigma = \sqrt{Var(X)}$

En statistiques, deux paramètres sont très importants : la moyenne (aussi appelée espérance) et la variance. La variance permet d’exprimer la variation d’une variable autour de la moyenne. L’écart type est la racine carrée de la variance.

Probabilité d'une variable aléatoire

- Soit P la probabilité d'obtenir des valeurs données avec la variable aléatoire X.
- Pour une variable continue:
  - *P*(X=x) = 0 où x est une valeur fixe donnée
  - P(-∞<X<+∞) = 1
- Fonction de répartition: F(x) = *P*(X<x)
- Propriété: Si X suit une loi normale de moyenne μ et écart type σ :

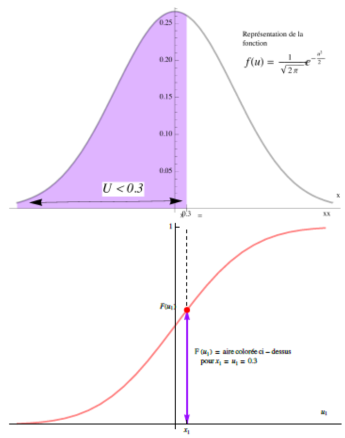
$$P(X < x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} exp^{\frac{-(z-\mu)^2}{2\sigma^2}} dz$$

Considérons la probabilité qu'une variable aléatoire X continue prenne une valeur donnée. Pour toute valeur exacte x, cette probabilité est nulle, car il y a une infinité de valeurs possibles. Par contre, il est certain que toute valeur de X sera entre -∞ et +∞, donc sa probabilité d'appartenir à cet intervalle est égale à 1. Comme on ne peut considérer des valeurs exactes, on s'intéresse à la probabilité que X soit dans un intervalle. On s'intéresse en particulier à la probabilité pour que X soit inférieur à une valeur x donnée. Cette probabilité qui dépend de x s'appelle fonction de répartition. La fonction de répartition d'une variable dite « normale » est connue et donnée par l'équation sur la diapositive. Elle s'exprime en fonction de sa moyenne et de son écart type.

Loi normale centrée réduite

$$F(x) = P(X < x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} exp^{\frac{-u^2}{2}} du$$

- Soit X une variable aléatoire normale de moyenne μ et écart type σ
- Montrer que $U = \frac{X-\mu}{\sigma}$ est normale centrée réduite



La loi normale centrée réduite est une fonction de répartition normale dont la moyenne est nulle et l'écart type de 1. On peut déduire toute loi normale à partir des valeurs de cette loi par simple changement de variable. Montrer qu'en faisant le changement de variable proposé sur la diapositive pour une loi normale quelconque, on retrouve bien une loi normale centrée réduite. La diapositive qui suit donne la réponse.

Démonstration

$u = \frac{z-\mu}{\sigma}$, d'où $du = \frac{dz}{\sigma}$

$$P(Z < z_1) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} dz$$

avec $u = \frac{z-\mu}{\sigma}$, $u_1 = \frac{z_1-\mu}{\sigma}$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{u_1} e^{-\frac{1}{2}(u)^2} (\sigma\, du)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u_1} e^{-\frac{1}{2}(u)^2} du$$

$$= P(U < u_1)$$

Exercices

- Montrer que :

1. $P(X > x) = 1 - F(x)$
2. $P(x_1 < X < x_2) = F(x_2) - F(x_1)$
3. Soit U une variable aléatoire normale centrée réduite :
   $P(U < -u) = 1 - F(u)$

Les solutions sont sur la diapositive suivante.

Corrections

- Démonstrations:

1. $P(X > x) = 1 – F(x)$

$P(-\infty<X<+\infty) = 1$

Or $P(X > x) + P(X < x) = P(-\infty<X<+\infty)$

Donc $P(X > x) = 1 – P(X < x)$

2. $P(x_1 < X < x_2) = F(x_2) – F(x_1)$

$P(X < x_2) = P(x_1 < X < x_2) + P(X < x_1)$

Donc $P(x_1 < X < x_2) = P(X < x_2) - P(X < x_1)$

Corrections

- Démonstration:

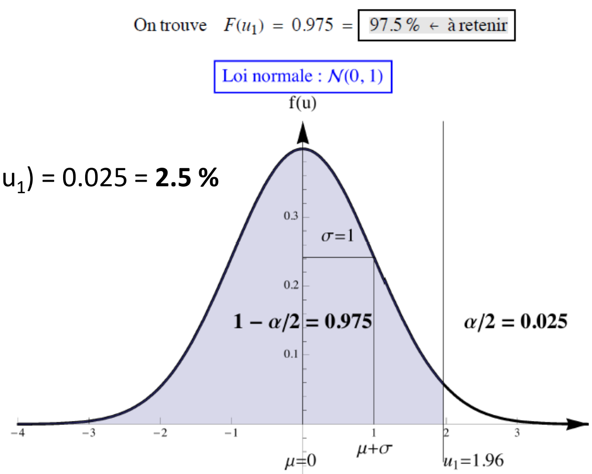
3. Soit U une variable aléatoire normale centrée réduite :
   $P(U < -u) = 1 - F(u)$

$P(U < -u) = P(U > u)$ car la loi normale est paire

Donc : $P(U < -u) = P(U > u) = 1 - F(u)$ d'après la démonstration précédente

Valeur remarquable

Pour $\mathbf{u_1 = 1.96}$: $P(U > u_1) = 0.025 =$ **2.5 %**



Il existe des valeurs remarquables liées aux variables suivant la loi normale centrée réduite. Nous nous intéresserons ici en particulier à la valeur u_1 telle que la probabilité pour qu'une telle variable soit supérieure à cette valeur u_1 vaut 0.025 (ie 2.5% de chances).

Exercice

- Soit U une variable aléatoire normale centrée réduite, pour quelle valeur *u1* a-t-on $P(|U| < u_1) = 0.95 = 95\%$?

Voir la correction sur la diapositive suivante.

Correction

$P(|U| < u) = P(-u < U < u) =$
$P(U < u) - P(U < -u) =$
$P(U < u) - P(U > u) =$
$P(U < u) - (1 - P(U < u))$
$= 2\,P(U < u) - 1$

$P(U < -u) = P(U > u)$ *car la loi normale centrée réduite est une fonction paire*

Donc

$$P(|U| < u_1) = 0.95$$
$$\Leftrightarrow 2P(U < u_1) - 1 = 0.95$$
$$\Leftrightarrow 2P(U < u_1) = 1.95$$
$$\Leftrightarrow P(U < u_1) = 0.975$$

$u_1 = 1.96$

Théorême Central-Limite

- Soient n variables aléatoire indépendantes $X_1, X_2, \ldots, X_n$ suivant toutes la même loi X de moyenne $E(X) = \mu$ et de variance $Var(X) = \sigma^2$
- E.g. $X_1, X_2, \ldots, X_n$ peuvent être plusieurs mesures d'un même phénomène (échantillon)
- La moyenne $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ suit **une loi normale** de moyenne μ et de variance $\frac{\sigma^2}{n}$ quand n tend vers ∞ (en pratique $n > 30$)

Un théorème très connu en statistique stipule que si on considère plusieurs mesures d'un même phénomène, par exemple les mesures faites sur les différents individus d'un échantillon utilisé pour un test, la moyenne de leurs mesures suivra une loi normale si l'échantillon est suffisamment grand.

Intervalle de variation

- Soit X une variable aléatoire de moyenne μ et écart type σ
- Montrer que sa moyenne sur un grand échantillon d'effectif **n** a 95% de chance d'appartenir à l'intervalle : $\left[\mu - 1.96 \times \frac{\sigma}{\sqrt{n}}\ ,\ \mu + 1.96 \times \frac{\sigma}{\sqrt{n}}\right]$

En utilisant le théorème central-limite vu précédemment, montrez ce qui est demandé sur la diapositive. La correction se trouve sur la diapositive suivante.

Démonstration

La variable $U = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ suit une loi normale centrée réduite

Donc

$P(-1.96 < U < 1.96) = P(|U| < 1.96) = 0.95$

$\Leftrightarrow P\left(-1.96 \times \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < 1.96 \times \frac{\sigma}{\sqrt{n}}\right) = 0.95$

$\Leftrightarrow P\left(\mu - 1.96 \times \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + 1.96 \times \frac{\sigma}{\sqrt{n}}\right) = 0.95$

Intervalle de confiance

- Intervalle où peut se situer la moyenne réelle de la population entière connaissant la moyenne de l'échantillon

$$P\left(\mu - 1.96 \times \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + 1.96 \times \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$\Leftrightarrow P\left(\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$IC = \left[\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}\right]$$

- Pour les grands échantillons, on peut estimer σ comme la variance observée $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}$

De ces considérations, on peut définir un intervalle de confiance, c'est-à-dire l'intervalle où peut se trouver la moyenne réelle d'une population avec une probabilité de 95% lorsqu'on connaît la moyenne d'un échantillon.

Raisonnement inférentiel

- Quand on prélève un échantillon d'une population, sa moyenne est une variable aléatoire (il y a autant de valeur possible pour la moyenne que d'échantillons)
- On veut savoir si une valeur hypothétique μ_0 est une moyenne possible de la population avec un risque d'erreur de moins de 5%
- On pose 2 hypothèses:
  - $H_0 : \mu = \mu_0$
  - $H_1 : \mu \neq \mu_0$
- On cherche à réfuter H_0 avec moins de 5% de risque

Le raisonnement inférentiel va procéder par l'absurde. On pose une hypothèse sur une valeur comme étant la moyenne d'une mesure sur la population générale. On sélectionne un échantillon d'individus sur lesquels on effectue les mesures. On dispose ainsi d'un intervalle de confiance centré autour de la moyenne des mesures réalisées sur l'échantillon. On cherche alors à montrer que la valeur hypothétique ne peut pas être la moyenne car elle est en dehors de l'intervalle de confiance. Elle a donc moins de 5% de chance d'être la vrai valeur moyenne.
La diapositive présente et la suivante résume le raisonnement.

Raisonnement inférentiel

- On calcule l'intervalle de confiance:

$$IC = \left[\bar{X} - 1.96 \times \frac{\hat{\sigma}}{\sqrt{n}}, \bar{X} + 1.96 \times \frac{\hat{\sigma}}{\sqrt{n}}\right]$$

- Si μ_0 est dans cet intervalle, on ne peut pas conclure
- Si μ_0 est hors de cet intervalle, on peut conclure que l'hypothèse H_0 est fausse avec un risque de 5%
- On dit alors que la moyenne de la population est **"significativement"** différente de μ_0

Intérêt du raisonnement inférentiel

- Il permet de généraliser un résultat sur toute la population à partir d'un échantillon
- En IHM : Si on fait un test sur un échantillon d'utilisateurs représentatifs, on peut conclure sur l'ensemble des utilisateurs possibles.

Exercice

- Le tabagisme actif se traduit par une teneur en cotinine dans le sang de plus de 200 μg/l
- Un fabricant de cigarettes électroniques prétend que ses cigarettes ne sont pas néfastes pour la santé.
- Le ministère de la santé décide de faire un test sur 100 personnes
- Pour ce test, le niveau moyen de cotinine est de 172.5 μg/l avec un écart type observé de 119.5 μg/l
- Ces nouvelles cigarettes sont-elles réellement sans danger pour la santé ?

La correction est sur la diapositive suivante.

Correction

- Hypothèses:
  - $H_0 : \mu = \mu_0$, la moyenne de l'échantillon n'est pas significativement différente de la valeur du tabagisme actif
  - $H_1 : \mu \neq \mu_0$, la moyenne de l'échantillon est significativement différente de la valeur du tabagisme actif
- Intervalle de confiance :

$$\left[172.5 - 1.96 \times \frac{119.5}{\sqrt{100}}, 172.5 + 1.96 \times \frac{119.5}{\sqrt{100}}\right] = [149.08, 195.92]$$

- Cet IC ne contient pas la valeur correspondant au tabagisme actif
- Donc, on peut rejeter l'hypothèse que les cigarettes sont nocives avec un risque de 5%
- PS: Dans la réalité, un risque de 5% n'est pas acceptable en santé (le risque acceptable est plutôt de l'ordre de 0.0001%)