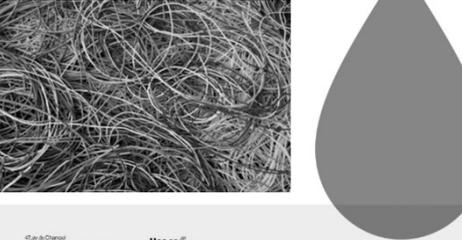


Soins infirmiers

Recherche 3 en science infirmière

Introduction aux tests statistiques et aux tailles d'effets

Pr Swann Pichon - Haute Ecole de Santé de Genève HES-SO
 Chercheur associé Université de Genève FPSE - Campus Biotech



09 Janvier 2023

H E S S
 Haute Ecole de Santé de Genève
 HES
 Haute Ecole de Santé de Genève
 CH de Charité
 1205 Genève
 +41 22 338 54 11
 info@hospreg.ch
 www.hospreg.ch
 HES-SO
 Université de Genève

Résumé des objectifs du dernier cours sur la démarche expérimentale

- Pouvoir situer les origines historiques de la méthode expérimentale, et les origines de la méthode statistique qui y est associée
- Pouvoir expliquer la différence **fondamentale** entre un lien de corrélation et un lien de causalité entre deux variables, et en quoi la démarche expérimentale permet de les différencier en mettant à l'épreuve l'hypothèse d'un lien de causalité entre les variables manipulées (VI) et la variable mesurée (VD)
- Contraster les avantages et inconvénients de la démarche corrélationnelle et de la démarche expérimentale
- Expliquer ce qui permet, dans la méthode expérimentale, de réduire les différences potentielles entre les participants et les groupes comparés
- Expliquer la notion d'effet principal et d'interaction entre deux variables indépendantes que le devis expérimental fait varier
- Identifier les différents types de devis expérimentaux et les deux grand types de mesures (inter versus intra-sujets)

HEDS-Genève 2

Plan et objectifs pédagogiques du cours

L'objectif pédagogique général du cours est de donner **une vue d'ensemble** de la notion de test statistique et des hypothèses qu'ils permettent de tester.

Nb: Notez que des notions comme les conditions d'applicabilité du test, le calcul et la distribution de la statistique, ou encore les notions de tests paramétriques et non-paramétriques sont des notions importantes non abordées dans ce cours

- Pouvoir expliquer la logique globale d'un test d'hypothèse, sans mathématiques
- Pouvoir expliquer le but d'un test d'association (encore appelé test de corrélation) et le but d'un test de différence de moyennes
 - Association: Comprendre la logique du test du χ^2 (Chi2)
 - Association: Comprendre la logique du test de **corrélation linéaire**
 - Association: Comprendre la logique du test de **régression linéaire**
 - Différence de moyennes : Comprendre la logique du **test de Student** et de l'**ANOVA**
- Pouvoir expliquer la notion de taille d'effet et son lien par rapport à la pertinence clinique d'une intervention

HEDS-Genève 3

Un premier mot avant de commencer

- L'objectif de ce cours était de vous faire comprendre la logique des principaux test statistiques et les inférences qu'elles permettent de tester
- Nous ne nous attendons pas à ce que vous reteniez tous les détails présentés aujourd'hui, ni à savoir appliquer ces tests concrètement, mais au moins à les lire et les comprendre
- **Nous vous encourageons à poser toutes les questions qui pourront vous aider et aider vos camarades à mieux comprendre les objectifs abordés dans ce cours**

HEDS-Genève 4

Les statistiques descriptives et différentielles - Définitions

- La statistique peut se définir comme l'art de collecter, d'analyser, de visualiser et d'interpréter des « données » pour évaluer la « fiabilité » des décisions fondées sur ces données (i.e. existe-t-il un lien entre deux variables observées? ou une différence de moyenne entre la valeur d'une variable mesurée dans le cadre de 2 conditions d'une étude expérimentale, ou une différence entre deux échantillons observés?). On distingue deux types de statistiques complémentaires
- Les statistiques descriptives
 - Sont des méthodes pour **décrire** la tendance centrale (moyenne, médiane) d'une variable, sa fréquence, sa dispersion (déviatoin standard, IC95% ...)
 - **Représenter** ces tendances et ces distributions via des graphiques (boites à moustache, histogrammes, graphiques à barres, nuages de points ...)
- Les statistiques inférentielles (aka les tests)
 - Sont des méthodes probabilistes utilisées (par ex.) pour estimer l'existence de corrélations ou de différences de moyennes entre deux variables
 - Ces tests sont utilisés car on a jamais accès à une population dans son ensemble, mais toujours à un sous-échantillon. Le but des tests statistiques est d'étendre les propriétés de l'échantillon à celui de la population duquel il est issu.
 - Ces tests sont faits en acceptant une certaine probabilité d'erreur => la valeur de p (probabilité de réaliser un faux positif)

HEDS-Genève 5

Statistiques descriptives et statistiques inférentielles

Echelle de mesure	Décrite : tendance centrale	Illustration : graphique	de différences entre moyennes +2 moyennes	Tests de différences entre +2 moyennes
Catégorielle	Mode %, effectif	Carte ou fréquence	appariés NA: Chi-carré/Chi-square (χ^2) A: Cochran (Q)	NA = non-appariés A = appariés
Ordinale	Médiane %, effectif	Boite à moustache	Test de Wilcoxon (U)	NA: Kruskal-Wallis (H) A: Friedman (Q)
Continue	Moyenne	Boite à moustache	Test T de Student (t)	NA: Analyse de variance/ Analysis of variance (ANOVA) A: ANOVA à mesures répétées

HEDS-Genève Repeated 6

Statistiques descriptives

Type de Variable	Décrire la tendance centrale dans un texte (dispersion)	Illustrer la tendance par un graphique
Catégorielle (sex: 0:10, 1:15)	% et effectif	i.e. Camembert ou barres de fréquences

Boîtes à moustache HEDS-Genève 7

Illustrer une médiane et les valeurs extrêmes d'une distribution à l'aide d'une boîte à moustache (BAM)

- Une BAM permet de visualiser en un seul graphique la médiane, les quartiles et les valeurs extrêmes de l'échantillon
- Souvent utilisée pour illustrer des données discrètes mais aussi continues car pratique pour visualiser les valeurs extrêmes (outliers)
 - Quartiles: Les 3 valeurs (Q1, Q2, Q3) qui divisent un jeu de données en 4 jeux de même taille
 - Médiane: la valeur Q2 qui divise un jeu de données en 2 jeux de même taille
 - Distance interquartile (IQR) = Q3 - Q1
 - Les valeurs extrêmes d'une BAM sont $< (Q1 - 1.5 \cdot IQR)$ ou $> (Q3 + 1.5 \cdot IQR)$

Distribution de l'échantillon (n=100)
 Distribution non problématique (n=100)
 Autre distribution possiblement problématique HEDS-Genève 8

Illustrer une moyenne et l'intervalle de confiance liée à son estimation

- Lorsque l'on cherche à estimer une moyenne au sein d'un échantillon, il est important de l'accompagner d'un indicateur de dispersion comme l'intervalle de confiance à 95% (IC95) qui renseigne sur la précision de son estimation
- L'intervalle de confiance à 95% est l'intervalle de valeur qui a 95% de chance de contenir la vraie valeur de la moyenne dans la population cible

- Exemple1: La réduction du poids suite au régime A est de -0.5 Kg [-1.2 +1.1]
- Exemple2: La réduction du poids suite au régime B est de -1.6 Kg [-2.3 -1.3]

HEDS-Genève 9

Statistiques inférentielles – Tests d’association et Tests de différences de moyennes

Type de Variable	Tests d’association (paramètre du test)		Tests de différence de moyennes (paramètre du test) NA = non-appariés; A = appariés	
	Lien ou corrélation entre 2 variables	Régression entre 2 variables	Entre 2 moyennes	Entre +2 moyennes
Catégorielle (sexe: ♀;10, ♂;15)	Test Chi-carré (χ^2) Test Phi (ϕ) Test exact de Fisher (F)	Régression logistique (OR \Rightarrow odds ratio; RR \Rightarrow risque relatif)	NA: Chi-carré (χ^2) A: Chi-carré de McNemar (χ^2)	NA: Chi-carré (χ^2) A: Cochran (Q)
Discrète (note: [1,2,3,4,5,6])	Rho Spearman (ρ) Tau de Kendall (τ)	Régression linéaire (R ou R^2 et b ou β)	NA et A: Test de Mann-Whitney-Wilcoxon (U)	NA: Kruskal-Wallis (H) A: Friedman (Q)
Continue (poids: 70,3 kg)	Corrélation de Pearson (r)		NA et A: Test T de Student (t)	NA: ANOVA (F) A: ANOVA à mesures répétées (F)

HEDS-Genève 10

Grille des tests statistiques ordonnés par type de groupes comparés (dépendant versus indépendant)

Groupes indépendants = Groupes non-appariés (NA) = individus différents dans chaque groupe
Groupes dépendants = Groupes appariés (A) = mêmes individus mesurés dans chaque groupe

Level of measurement of dependent variable	Group Comparisons: Number of groups (the independent variable)				Correlational analyses (To examine relationship strength)
	2 Groups		3+ Groups		
	Independent Groups Tests	Dependent Groups Tests	Independent Groups Tests	Dependent Groups Tests	
Nominal (Categorical)	χ^2 p. 401 (or Fisher's exact test) p. 402	McNemar's test p. 402	χ^2 p. 401	Cochran's Q	Phi coefficient (dichotomous) or Cramer's V (not restricted to dichotomous) p. 403
Ordinal (Rank)	Mann-Whitney Test p. 396	Wilcoxon signed ranks test p. 396	Kruskal-Wallis H test p. 400	Friedman's test p. 400	Spearman's rho (or Kendall's tau) pp. 403
Interval or Ratio (Continuous)*	Independent group t test pp. 394-395	Paired t test p. 396	ANOVA pp. 396-399	RM-ANOVA pp. 400	Pearson's r p. 402
	Multifactor ANOVA for 2+ independent variables p. 398				
	RM-ANOVA for 2+ groups x 2+ measurements over time p. 424				

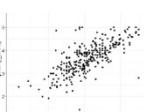
*For distributions that are markedly nonnormal or samples that are small, the nonparametric tests in the row above (for ordinal measures) may be needed.

HEDS-Genève 11

Quatre grandes familles de tests (le cas des variables continues a été considéré pour l'exemple)

Tests de corrélation (lien)	Tests de régression (lien)	Tests de différences entre 2 moyennes	Tests de différences entre +2 moyennes
-----------------------------	----------------------------	---------------------------------------	--

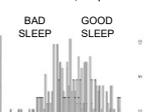
- Ce set de données représente les mesures d’humeur selon la qualité du sommeil chez ~200 hommes et femmes. Les 4 tests suivants et les graphiques associés ont été employé sur les mêmes données
- Corrélation : la variable humeur et qualité du sommeil varient-elles ensemble? (oui, positivement)
- Régression : existe-t-il une droite de pente non nulle qui approxime cette relation (oui, si la pente était plate, il n’y aurait pas de lien entre les variables).
- T-test : existe-t-il une différence de moyenne entre l’humeur des bons et des mauvais dormeurs? (oui)
- ANOVA : existe-t-il une différence de moyenne entre l’humeur des bons, moyens et mauvais dormeurs? (oui)



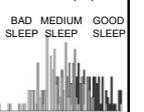
Corrélation de Pearson (r)



Régression linéaire



Student t-test



ANOVA

HEDS-Genève 12

Principe général d'un test statistique

- Un test statistique est une procédure de décision entre deux hypothèses
 - H_0 : l'hypothèse nulle (pas d'effet)
 - H_1 : l'hypothèse alternative (un effet)
- Le principe de chaque test statistique consiste à
 - Calculer un score statistique qui «résume les données testées»
 - Puis décider si ce score penche plutôt en faveur de l'hypothèse nulle ou de l'hypothèse alternative, en acceptant un risque d'erreur de se tromper (la valeur p , qui exprime la probabilité de réaliser un faux positif)
- Chaque statistique estime un score particulier
 - Score χ^2 lorsque l'on traite les fréquences d'un tableau de répartition entre deux variables
 - Score r de Pearson lorsque l'on teste une corrélation entre deux variables continues
 - Score t de Student si on teste une différence entre 2 moyennes
 - Score F de l'ANOVA si on teste une différence entre +2 moyennes

HEDS-Genève 13

La p-valeur késako

- La valeur de p est une probabilité – elle varie entre 0 et 1 – qui quantifie la probabilité de se tromper lorsque l'on déclare un test significatif (i.e. que l'on rejette H_0)
 - La valeur de p détermine la significativité d'un test
 - Par convention, la valeur de p doit être inférieure à 5% pour que le test soit déclaré significatif
- Par exemple, un test de corrélation entre la variable «qualité du sommeil» et le «niveau d'humeur» retourne une valeur $r = 0.73$, avec une valeur de $p = 0.001$.
- Chaque score statistique a une distribution particulière attendue sous H_0 , qui est fonction du nombre d'observations dans l'échantillon.

HEDS-Genève 14

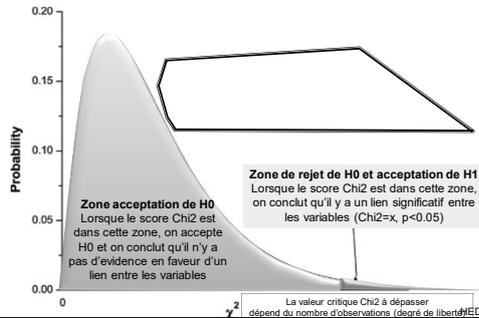
La p-valeur expliquée à mon chat

P-valeur ou je fais un malheur !

HEDS-Genève 15

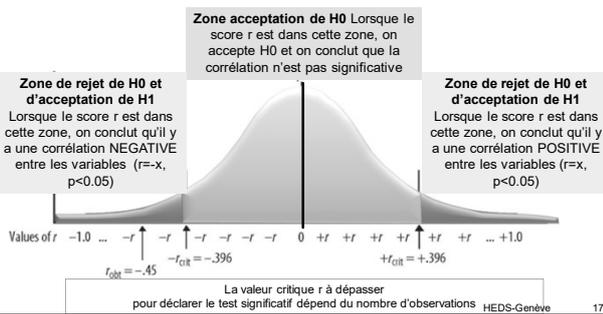
Exemple de Distribution du Chi-carré sous H0 – test d'indépendance du Chi2

- Cette distribution, uniquement positive, permet de tester l'hypothèse d'un lien entre deux variables (ex: facteur de risque X patients)



Exemple de Distribution du score r de Pearson sous H0 - Test bilatéral de corrélation de Pearson

- Cette distribution, centrée sur 0, permet de tester l'hypothèse qu'un score r positif ou négatif diffère ou non de 0



Tests d'hypothèses et inférences faites selon les tests (Chi2, corrélation, régression, T-test, ANOVA)

- **H0 = Hypothèse nulle** ~ évidence insuffisante en faveur d'un effet ($p \geq 0.05$)
- **Chi2 (χ^2):** Le facteur d'exposition n'a pas d'effet significatif sur l'outcome clinique
- **Corrélation:** le coefficient de corrélation n'est pas significativement différent de 0
- **Régression:** la pente (ou l'intercept) de la droite n'est pas significativement différente de 0
- **T-Test de Student:** La différence entre deux moyennes n'est pas signif. différente de 0
- **ANOVA:** Les n moyennes comparées ne sont pas signif. différentes les unes des autres
- **H1 = Hypothèse alternative** ~ évidence en faveur d'un effet ($p < 0.05$)
- **Chi2:** Le facteur d'exposition est lié significativement à l'outcome clinique
- **Corrélation:** le coefficient de corrélation est signif. différent de 0
- **Régression:** la pente (ou l'intercept) de la droite est signif. différente de 0
- **T-Test de Student:** Les deux moyennes comparées sont signif. différentes
- **ANOVA:** Au moins une des n moyennes comparées est signif. différente des autres

On distingue deux types d'erreurs en statistique, ainsi que deux types de risques

- L'erreur de **faux positif** consiste à déclarer un test significatif (je déclare qu'un non-patient est malade) alors qu'en fait il ne l'est pas réellement (erreur de commission)
- Le **risque alpha (α)**, désigne le risque de déclarer un **faux positif** lorsque l'on rejette H_0 . C'est ce seuil alpha que l'on fixe par convention à 5% et que l'on utilise pour décider si un test est significatif ou non via la **valeur p**.
- Le risque alpha est lié à la **spécificité d'un test/diagnostic**, en favorisant la détection des patients vraiment malades, c'est-à-dire en limitant le nombre de faux positifs
- L'erreur de **faux négatif** consiste à ne pas déclarer un test significatif (je déclare qu'un vrai patient n'est pas malade) alors qu'en fait il l'est réellement (erreur d'omission)
- Le **risque beta**, généralement de 20%, désigne le risque de faire un faux négatif (omission). Le risque beta est important pour réaliser un **calcul de puissance statistique** qui vise à déterminer la taille minimale de l'échantillon que l'on doit collecter pour se donner une chance de 80% de détecter un effet s'il existe.
- Le risque beta est lié à la **sensibilité d'un test / diagnostic**, en favorisant la détection de patients véritablement malade, c'est-à-dire en minimisant au mieux le nombre de faux négatifs.

Diagnostic

		Etat réel du patient	
Diagnostic		Vrai Positif Détectations Correctes	Faux Positif Commissions α
		Faux négatif Omissions β	Vrai Négatif Rejections Correctes

HEDS-Genève 19

Plan et objectifs pédagogiques du cours

L'objectif pédagogique général du cours est de donner **une vue d'ensemble** de la notion de test statistique et des hypothèses qu'ils permettent de tester.

Nb: Notez que des notions comme les conditions d'applicabilité du test, le calcul et la distribution de la statistique, ou encore les notions de tests paramétriques et non-paramétriques sont des notions importantes non abordées dans ce cours

- Pouvoir expliquer la logique globale d'un test d'hypothèse, sans mathématiques
- Pouvoir expliquer le but d'un test d'association (encore appelé test de corrélation) et le but d'un test de différence de moyennes
 1. Association: Comprendre la logique du test du χ^2 (**Chi2**)
 2. Association: Comprendre la logique du test de **corrélation linéaire**
 3. Association: Comprendre la logique du test de **régression linéaire**
 4. Différence de moyennes : Comprendre la logique du **test de Student** et de l'**ANOVA**
- Pouvoir expliquer la notion de taille d'effet et son lien par rapport à la pertinence clinique d'une intervention

HEDS-Genève 20

Quatre grandes familles de tests statistiques – Apprenons à distinguer les 4 tests pour les variables continues

Type de Variable	Tests d'association (paramètre du test)		Tests de différence de moyennes (paramètre du test) <small>NA = non-appariés; A = appariés</small>	
	Lien ou corrélation entre 2 variables	Régression entre 2 variables	Entre 2 moyennes	Entre +2 moyennes
Catégorielle <small>(sex: 0,10, 3,15)</small>	Test Chi-carré (χ^2) Test PHI (ϕ) Test exact de Fisher (F)	Régression logistique (OR =>odds ratio; RR=>risque relatif)	NA: Chi-carré (χ^2) A: Chi-carré de McNemar (χ^2)	NA: Chi-carré (χ^2) A: Cochran (Q)
Discrete <small>(note: 1,2,3,4,5,6)</small>	Rho Spearman (ρ) Tau de Kendall (τ)	Régression linéaire (R ou R² et b ou β)	NA et A: Test de Mann-Whitney-Wilcoxon (U)	NA: Kruskal-Wallis (H) A: Friedman (Q)
Continue <small>(poids: 70,3 kg)</small>	Corrélation de Pearson (r)		NA et A: Test T de Student (t)	NA: ANOVA (F) A: ANOVA à mesures répétées (F)

HEDS-Genève 21

Le principe du test Chi² (Chi-square, χ^2)

- De nombreuses études observationnelles recueillent des données de variables catégorielles et résument les effectifs de ces catégories dans un **tableau de répartition des effectifs (ou de contingence)**.
- Par exemple, en épidémiologie, le test du Chi² est souvent utilisé dans les études de cas contrôle pour tester l'hypothèse d'un lien entre un facteur d'exposition (exposé/non-exposé) et un facteur de santé (ex: malade/non malade)
- La statistique du test du Chi² sert à tester l'hypothèse qu'il existe une association entre deux variables catégorielles.
 - H0: Les deux facteurs catégoriels sont indépendants (pas de relation entre les facteurs)
 - H1: Les deux facteurs sont dépendants (relation entre les facteurs)
- La **logique du test** vise à comparer les effectifs du tableau de répartition des cas observés, aux effectifs théoriques que l'on devrait observer s'il n'existait pas de relation entre les variables catégorielles.
- Énoncé autrement, ce test vise à mettre en évidence un pattern d'interaction entre deux variables catégorielles

	Effectifs observés (nombre de cas)		Effectifs théoriques que l'on devrait observer s'il n'y avait pas de relation (H0)	
	Malades (Cases)	Non-malades (Controls)	Malades (Cases)	Non-malades (Controls)
Exposed	A	B	A _{théo.}	B _{théo.}
Non-exposed	C	D	C _{théo.}	D _{théo.}

22

Un exemple récent – le statut tabagique est-il lié à la sévérité de la COVID-19?

TABLE 1. Reported outcomes among COVID-19 patients of all ages, by hospitalization status, underlying health condition, and risk factor for severe outcome from respiratory infection — United States, February 12–March 28, 2020

Underlying health condition/Risk factor for severe outcomes from respiratory infection (no., % with condition)	No. (%)			
	Not hospitalized	Hospitalized, non-ICU	ICU admission	Hospitalization status unknown
Total with case report form (N = 74,439)	12,217	5,285	1,069	55,868
Missing or ambiguous status for all conditions (67,277)	2,824	4,268	814	55,843
Total with completed information (7,162)	5,143	1,427	457	525
One or more conditions (2,692, 37.6%)	1,388 (27)	732 (71)	358 (78)	214 (41)
Diabetes mellitus (784, 10.9%)	331 (6)	251 (24)	148 (32)	54 (10)
Chronic lung disease* (656, 9.3%)	360 (7)	152 (15)	96 (21)	47 (9)
Cardiovascular disease (647, 9.0%)	239 (5)	242 (23)	132 (29)	34 (6)
Immunocompromised condition (264, 3.7%)	141 (3)	63 (6)	41 (9)	19 (4)
Chronic renal disease (213, 3.0%)	51 (1)	95 (9)	56 (12)	11 (2)
Pregnancy (143, 2.0%)	72 (1)	31 (3)	4 (1)	36 (7)
Neurologic disorders, neurodevelopmental, intellectual disability (52, 0.7%) [†]	17 (0.3)	25 (2)	7 (2)	3 (1)
Chronic liver disease (41, 0.6%)	24 (1)	9 (1)	7 (2)	1 (0.2)
Other chronic disease (1,182, 16.5%) [‡]	583 (11)	359 (35)	170 (37)	70 (13)
Former smoker (165, 2.3%)	97 (2)	47 (4)	3 (1)	7 (1)
Current smoker (96, 1.3%)	61 (1)	27 (2)	5 (1)	8 (2)
None of the above conditions [§] (4,470, 62.4%)	3,755 (73)	309 (29)	99 (22)	311 (59)

Abbreviations: ICU = intensive care unit.

* Includes any of the following: asthma, chronic obstructive pulmonary disease, and emphysema.

[†] For neurologic disorders, neurodevelopmental, and intellectual disability, the following information was specified: dementia, memory loss, or Alzheimer's disease (17); seizure disorder (5); Parkinson's disease (4); migraine/headache (4); stroke (2); autism (2); aneurysm (2); multiple sclerosis (2); neuropathy (2); hereditary spastic paraplegia (1); myasthenia gravis (1); intracranial hemorrhage (1); and altered mental status (1).

[‡] For other chronic disease, the following information was specified: hypertension (113); thyroid disease (37); gastrointestinal disorder (32); hyperlipidemia (29); cancer or history of cancer (29); rheumatoid disorder (10); human immunodeficiency virus (11); arthritis (17); arthritis, nonrheumatoid (including non-rheumatoid spondylitis) (14);

Source: <https://www.cdc.gov/mmwr/volumes/69/wr/mm6913e2.htm#T1>

23

Un exemple récent – le statut tabagique est-il lié à la sévérité de la COVID-19?

	Effectifs observés (nombre de cas)			Effectifs théoriques que l'on devrait observer s'il n'y avait pas de relation (H0)		
	Severe COVID	Mild COVID	Effectif Total	Severe COVID	Mild COVID	Effectif Total
Smokers / past smokers	105	141	246	55	191	246
Non-smokers	1389	5002	6391	1439	4952	6391
% total	22.5%	77.5%		22.5%	77.5%	

	% observés			% théoriques s'il n'y avait pas de relation (H0)		
	Severe COVID	Mild COVID	Total	Severe COVID	Mild COVID	Total
Smokers / past smokers	43%	57%	100%	22.5%	77.5%	100%
Non-smokers	22%	78%	100%	22.5%	77.5%	100%

24

Tableau de contingence et le rapport des chances (odds ratio ou OR)

- Le rapport des chances (ou rapport des cotes, ou odds ratio (OR) en Anglais), est une autre mesure similaire de l'association (interaction) entre le facteur d'exposition et l'outcome clinique considéré.
- L'OR est le rapport entre la chance de survenue de la maladie dans le groupe exposé, et la chance de survenue de la maladie dans le groupe non-exposé.
- L'OR est la mesure d'association utilisée dans la régression logistique
- Moins intuitif à comprendre que le RR

$$OR = \frac{A/B}{C/D} = \frac{\text{A chances contre B de survenue de la maladie dans le groupe exposé}}{\text{C chances contre D de survenue de la maladie dans le groupe non-exposé}}$$

- L'OR est une taille d'effet comprise entre $[0, +\infty]$
 - Si $OR = 1$, il n'y a pas d'association entre les deux facteurs
 - Si $OR > 1$ (**facteur de risque**) le rapport de chances est OR fois **plus grand** chez le groupe exposé que non-exposé
 - Si $OR < 1$ (**facteur protecteur**) le rapport de chances est OR fois **moins grand** chez le groupe exposé que non-exposé
- Dans notre exemple, l'OR = 2.68 IC95% [2.07, 3.47]
- Lorsque l'**intervalle de confiance** d'un rapport de chances comprend la valeur 1, on sait d'avance que le test statistique ne pourra rejeter l'hypothèse nulle (pas d'évidence en faveur d'un lien).

Vidéo explicative – Taille d'effet RR et OR



Plan et objectifs pédagogiques du cours

L'objectif pédagogique général du cours est de donner **une vue d'ensemble** de la notion de test statistique et des hypothèses qu'ils permettent de tester.

Nb: Notez que des notions comme les conditions d'applicabilité du test, le calcul et la distribution de la statistique, ou encore les notions de tests paramétriques et non-paramétriques sont des notions importantes non abordées dans ce cours

- Pouvoir expliquer la logique globale d'un test d'hypothèse, sans mathématiques
- Pouvoir expliquer le but d'un test d'association (encore appelé test de corrélation) et le but d'un test de différence de moyennes
 - Association: Comprendre la logique du test du χ^2 (**Chi2**)
 - Association: Comprendre la logique du test de **corrélation linéaire**
 - Association: Comprendre la logique du test de **régression linéaire**
 - Différence de moyennes : Comprendre la logique du **test de Student** et de l'**ANOVA**
- Pouvoir expliquer la notion de taille d'effet et son lien par rapport à la pertinence clinique d'une intervention

Tests de corrélation et de régression

Type de Variable	Tests d'association (paramètre du test)		Tests de différence de moyennes (paramètre du test) NA = non-appariés; A = appariés	
	Lien ou corrélation entre 2 variables	Régression entre 2 variables	Entre 2 moyennes	Entre +2 moyennes
Catégorielle (sexe: ♀;10, ♂;15)	Test Chi-carré (χ^2) Test Phi (ϕ) Test exact de Fisher (F)	Régression logistique (OR =>odds ratio; RR=>risque relatif)	NA: Chi-carré (χ^2) A: Chi-carré de McNemar (χ^2)	NA: Chi-carré (χ^2) A: Cochran
Discrete (note: [1,2,3,4,5,6])	Rho Spearman (ρ) Tau de Kendall (τ)	Régression linéaire (R ou R^2 et b ou β)	NA et A: Test de Mann-Whitney-Wilcoxon (U)	NA: Kruskal-Wallis (H) A: Friedman (Q)
Continue (poids: 70,3 kg)	Corrélation de Pearson (r)		NA et A: Test T de Student (t)	NA: ANOVA A: ANOVA à mesures répétées (F) HEDS-Genève 31

Tester l'hypothèse d'une liaison entre deux variables

- 306 patients ont estimé la qualité de leur **sommeil** (1-8) et de leur **humeur** (1-5) durant le mois passé
- Ci-dessous, l'hypothèse d'une association entre sommeil et humeur a été testée de trois façons différentes sur les mêmes données.
- Conclusion: les trois tests donnent des résultats concordants, indiquant une association positive entre l'humeur et la qualité du sommeil

Corrélation (Pearson r)
VI et VD continues

Résultat: je sais que ces deux variables sont associées avec une force d'association qui vaut r

Régression linéaire
VI et VD continues

Résultat: je sais que ces deux variables sont associées et je dispose d'une équation linéaire ou logistique qui me permet de prédire la qualité du sommeil (VD) à partir de l'humeur (VI)

Régression logistique
VI catégorielle, VD continue

Résultat: je sais que ces deux variables sont associées et je dispose d'une équation linéaire ou logistique qui me permet de prédire la qualité du sommeil (VD) à partir de l'humeur (VI)

Tester l'hypothèse d'une liaison entre deux variables

- Si ces deux mêmes variables n'étaient pas associées - ce que nous pouvons faire en multipliant chaque observation par un chiffre aléatoirement tiré entre 0 et 1 - voici ce que nous pourrions observer pour chacun des tests
- Le nuage de point serait sphérique => pas de lien
- Les régressions auraient une pente nulle/plate => pas de lien

Corrélation (Pearson r)
VI et VD continues

Les deux variables ne sont pas corrélées

Régression linéaire
VI et VD continues

Les pentes des régressions ne sont pas significativement différentes de 0. Il n'y a donc pas d'association significative entre

Régression logistique
VI catégorielle, VD continue

Les pentes des régressions ne sont pas significativement différentes de 0. Il n'y a donc pas d'association significative entre

Corrélation de Pearson

- Une corrélation entre deux variables a pour objectif d'estimer la force et la direction de l'association entre deux variables via un coefficient de corrélation (r ou rho)
- Le test de corrélation vise à tester si ce coefficient est significativement différent de 0, ce qui dépendra de sa force et du nombre d'observation
- Le coefficient de corrélation de Pearson (r) est une mesure du degré et de la direction d'association entre deux variables. Il est compris entre [-1; 1]
 - Si r tend vers -1: corrélation négative
 - Si r tend vers +1: corrélation positive
 - Si r tend vers 0: corrélation nulle

Corrélation paramétrique (Pearson) = 0.73, p < 0.001
 Coefficient de corrélation non paramétrique (Spearman) = 0.74, p < 0.001

- Hypothèses du test de corrélation
 - H0: Le coefficient de corrélation r ne diffère pas de 0
 - H1: Le coefficient de corrélation diffère de 0

HEDS-Genève 34

La régression linéaire simple

- Une régression linéaire vise à estimer l'équation de la droite qui colle le mieux aux données, c'est-à-dire qui passe le plus près des points d'un nuage de données.
- Cette droite est un modèle mathématique simple qui permet de résumer les données.
- L'équation de cette droite est la droite qui minimise la distance entre les valeurs observées (en rouge), et les valeurs prédites par la droite (en blanc).
- L'équation d'une droite se définit à l'aide de deux paramètres :
 - L'ordonnée à l'origine (intercept en Anglais)** de la droite qui désigne la valeur de Y lorsque X = 0
 - La pente de la droite** (appelé aussi coefficient directeur) qui désigne son inclinaison. Si la pente tend vers 0, la droite est « plate » et le modèle ne prédit rien => il n'existe pas de relation linéaire / d'association entre les deux variables

Y = pente * X + ordonnée_à_l'origine + erreur

Ce qui est connu / observé
 Paramètres que l'on cherche à estimer

- Hypothèses pour le modèle linéaire dans son ensemble (F stat.)
 - H0: Le modèle estimé ne prédit pas mieux les données que le modèle nul
 - H1: Le modèle estimé prédit mieux les données que le modèle nul
- Hypothèses pour chacun des paramètres (t student stat.)
 - H0: Les paramètres de pente (resp. d'ordonnée_origine) ne diffèrent pas de 0
 - H1: Les paramètres de pente (resp. d'ordonnée_origine) diffèrent de 0

HEDS-Genève 35

Plan et objectifs pédagogiques du cours

L'objectif pédagogique général du cours est de donner une **vue d'ensemble** de la notion de test statistique et des hypothèses qu'ils permettent de tester.

Nb: Notez que des notions comme les conditions d'applicabilité du test, le calcul et la distribution de la statistique, ou encore les notions de tests paramétriques et non-paramétriques sont des notions importantes non abordées dans ce cours

- Pouvoir expliquer la logique globale d'un test d'hypothèse, sans mathématiques
- Pouvoir expliquer le but d'un test d'association (encore appelé test de corrélation) et le but d'un test de différence de moyennes
 - Association: Comprendre la logique du test du χ^2 (Chi2)
 - Association: Comprendre la logique du test de **corrélation linéaire**
 - Association: Comprendre la logique du test de **régression linéaire**
 - Différence de moyennes : Comprendre la logique du **test de Student** et de l'**ANOVA**
- Pouvoir expliquer la notion de taille d'effet et son lien par rapport à la pertinence clinique d'une intervention

HEDS-Genève 36

Tests de comparaison de moyennes

Type de Variable	Tests d'association (paramètre du test)		Tests de différence de moyennes (paramètre du test) NA = non-appariés; A = appariés	
	Lien ou corrélation entre 2 variables	Régression entre 2 variables	Entre 2 moyennes	Entre +2 moyennes
Catégorielle (sexe: ♀:10, ♂:15)	Test Chi-carré (χ^2) Test Phi (ϕ) Test exact de Fisher (F)	Régression logistique (OR => odds ratio; RR=>risque relatif)	NA: Chi-carré (χ^2) A: Chi-carré de McNemar (χ^2)	NA: Chi-carré (χ^2) A: Cochran (Q)
Discrete (note: [1,2,3,4,5,6])	Rho Spearman (ρ) Tau de Kendall (τ)	Régression linéaire (R ou R ² et b ou β)	NA et A: Test de Mann-Whitney-Wilcoxon (U)	NA: Kruskal-Wallis (H) A: Friedman (Q)
Continue (poids: 70,3 kg)	Corrélation de Pearson (r)		NA et A: Test T de Student (t)	NA: ANOVA A: ANOVA à mesures répétées (F)

HEDS-Genève 37

Tester l'hypothèse d'une différence de moyennes entre deux échantillons (VI catégorielle, VD continue)

- Le test-t de Student a pour but de tester l'hypothèse d'une différence entre les moyennes de deux échantillons
 - Si les échantillons sont non-appariés (mesures issues de groupes distincts, unpaired en Anglais): on utilise le t-test non-apparié
 - Si les échantillons sont appariés (mesures issues du même groupe, paired en Anglais): on utilise le t-test apparié
- Ex: On applique le test-t de Student pour tester l'hypothèse d'une différence entre les moyennes d'humeurs mesurées dans ces deux échantillons
 - H0: L'humeur ne diffère pas en fonction de la qualité du sommeil
 - H1: L'humeur diffère en fonction de la qualité du sommeil
- Résultats concordants avec les tests de corrélation ☺

T-test: Effet de la qualité du sommeil sur l'humeur
t(304)=13.1, p<0e-16***, Cohen's d=1.5

HEDS-Genève 38

La statistique de t du test de Student mesure une différence de moyenne normalisée par la variance des échantillons

$$t = \frac{\text{signal}}{\text{noise}} = \frac{\text{difference in means}}{\text{sample variability}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

\bar{x}_1, \bar{x}_2 = sample means
 s_1^2, s_2^2 = sample variances
 n_1, n_2 = sample sizes

HEDS-Genève 39

L'ANOVA pour tester l'hypothèse d'une différence de moyennes entre au moins trois échantillons

- Echantillons non-appariés (groupes distincts): on utilise l'ANOVA
- Echantillons appariés (issus du même groupe): ANOVA à mesures répétées
- On forme trois groupes d'individus selon la qualité de leur sommeil (mauvais, intermédiaire, bon)
- On pose la question de savoir si l'humeur moyenne de ces trois groupes diffèrent l'une de l'autre
 - H0: L'humeur moyenne entre les 3 groupes n'est pas significativement différente
 - H1: Au moins une moyenne diffère des autres
- L'ANOVA est significative – résultats concordants avec les autres test ☺
- Eta² représente la taille d'effet dans les ANOVA – elle est importante

Plan et objectifs pédagogiques du cours

L'objectif pédagogique général du cours est de donner une **vue d'ensemble** de la notion de test statistique et des hypothèses qu'ils permettent de tester.

Nb: Notez que des notions comme les conditions d'applicabilité du test, le calcul et la distribution de la statistique, ou encore les notions de tests paramétriques et non-paramétriques sont des notions importantes non abordées dans ce cours

- Pouvoir expliquer la logique globale d'un test d'hypothèse, sans mathématiques
- Pouvoir expliquer le but d'un test d'association (encore appelé test de corrélation) et le but d'un test de différence de moyennes
 - Association: Comprendre la logique du test du χ^2 (Chi2)
 - Association: Comprendre la logique du test de **corrélation linéaire**
 - Association: Comprendre la logique du test de **régression linéaire**
 - Différence de moyennes : Comprendre la logique du **test de Student** et de l'**ANOVA**
- Pouvoir expliquer la notion de taille d'effet et son lien par rapport à la pertinence clinique d'une intervention

La notion de taille d'effet (TE) quantifie la magnitude d'un effet, ce qui permet d'évaluer la pertinence clinique d'une intervention

- La **taille d'effet (TE)** désigne la force d'un phénomène, c'est-à-dire la force d'association entre deux variables ou la grandeur de la différence entre deux moyennes à l'aide d'un indicateur standardisé. Il existe différentes tailles d'effets
- La significativité statistique (p) ne donne pas d'information quant à la taille de l'effet considéré (ex: la p value est fonction de la taille d'effet et de l'effectif)
 - TE d'un lien entre variables catégorielles: RR ou OR
 - TE d'une corrélation: r de Pearson et rho de Spearman
 - TE d'un facteur de régression: R2 ou f2
- TE différence entre deux moyennes: d de Cohen
- TE d'un facteur d'ANOVA: Eta²

- Pourquoi est-ce important?
 - Parce qu'elle quantifie la force, et donc la pertinence clinique d'une intervention en fonction du gain escompté (et non en fonction de la significativité, qui ne dit rien sur la taille de l'effet). Introduire une nouvelle intervention prend du temps, des ressources etc. Il est donc important de pouvoir estimer le bénéfice réel pour le patient-e
 - Les tailles d'effets permettent de **conduire des méta-analyses** pour synthétiser la taille d'effet moyenne d'une famille d'intervention
 - Les tailles d'effets permettent enfin de **conduire des analyses de puissance** avant de débiter une expérience. Ces analyses visent à déterminer la taille de l'échantillon qu'il sera nécessaire de recruter pour être sûr de détecter un effet avec une taille d'effet donnée et une puissance donnée (le risque beta de déclarer un faux négatif, souvent fixé à 80%). Une étape obligatoire pour justifier un financement de projet de recherche

Les méthodologistes en science infirmière argumentent en faveur de toujours rapporter la taille d'effet dans un test statistique

J Nurs Educ. 2017 Nov;15(6):445-447. doi: 10.1016/j.nurse.2017.02.002.

The Purpose and Power of Reporting Effect Sizes in Nursing Education Research

Dorell Sautter, J
PMID: 29091202 DOI: 10.1016/j.nurse.2017.02.002

Abstract

Like many other fields and subfields within the social sciences, education, and medicine, nursing education research has a long history of giving p-values (associated with common statistical tests) a position of primary importance in interpreting study findings. Global, transdisciplinary efforts are underway to diminish the role p-values play in making judgments about the significance of study findings by changing the threshold p-value used to announce statistical significance, through the use of Bayesian statistics and **emphasizing effect sizes** (https://doi.org/10.1016/j.nurse.2017.02.002). **Keywords:** Effect sizes, p-values, statistical significance, Bayesian statistics, nursing education research. © Nurs Educ. 2017;15(6):445-447.

À l'instar de nombreux autres domaines et sous-domaines des sciences sociales, de l'éducation et de la médecine, la formation à la recherche en soins infirmiers accorde depuis longtemps aux valeurs p des tests statistiques une place de premier plan dans l'interprétation des résultats des études. Des efforts mondiaux et transdisciplinaires sont en cours pour diminuer le rôle que jouent les valeurs p dans les jugements sur l'importance des résultats des études, en modifiant le seuil de la valeur p utilisée pour prononcer la signification statistique (ex. p<0.01 plutôt que p<0.05), en utilisant les statistiques bayésiennes et en s'efforçant de signaler les tailles d'effet en même temps que les valeurs p. Dans le numéro de ce mois-ci, l'accent est mis sur les tailles d'effet et leur rôle dans l'amélioration de la valeur et de l'utilité des études de recherche sur la formation en soins infirmiers.

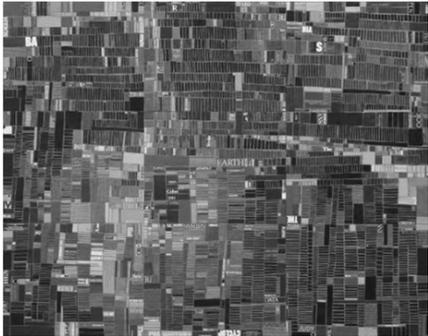
https://journals.healio.com/doi/10.3928/01484834-20171020-02 HEDS-Genève 46

Un dernier mot avant de finir

- L'objectif de ce cours était de vous faire comprendre la logique des principaux test statistiques et les inférences qu'elles permettent de tester
- Nous ne nous attendons pas à ce que vous reteniez tous les détails présentés aujourd'hui, ni à savoir appliquer ces tests concrètement, mais au moins à les lire et les comprendre
- Nous vous encourageons à poser toutes les questions qui pourront vous aider et aider vos camarades à mieux comprendre les objectifs abordés dans ce cours

HEDS-Genève 47

Merci de votre attention – et merci de poser toutes vos questions !



Imagined Time, Laura Frick
HEDS-Genève 48
