

Notions de base en statistique

Dr Cécile Couchoud

Démarche déductive

- « Vraie » valeur \rightarrow prévoir la valeur sur un échantillon futur $<$ fluctuations d'échantillonnage

- Intervalle de fluctuation

- ◆ Formule générale :

$$A \pm z_{\alpha/2} \sqrt{\text{var}A}$$

A: moyenne du paramètre

α : risque d'erreur que l'on consent (5%)

$z_{\alpha/2}$: donné par la table de la loi normale centrée réduite

$\sqrt{\text{var}A}$: variance du paramètre

Démarche inductive

- Valeur observée sur un échantillon : Estimation ponctuelle p_0
- Estimation par intervalle : intervalle de confiance

$$p_0 \pm z_{\alpha/2} \sqrt{\text{var } p_0}$$

- Interprétation : pour $\alpha=5\%$, $z_{\alpha/2} = 1.96$
« sur 100 échantillons successifs pris dans la population, 95 (en moyenne) conduisent à un intervalle de confiance qui contient le vrai pourcentage »

Quelques exemples de variances

- Prévalence :

$$\text{Var}(p_0) = p_0 \cdot (1-p_0) / n$$

- Taux d'incidence :

$$\text{Var}(\text{TI}) = m / PA^2$$

- Moyenne :

$$\text{Var}(m) = s^2 / n$$

- Odds ratio :

$$\text{Var}(\text{LnOR}) = 1/a + 1/b + 1/c + 1/d$$

Étapes d'un test statistique

Hypothèses à priori

- H_0 : hypothèse nulle
 - ◆ pas de vraie différence : ce qu'on observe est purement liée au hasard
 - ◆ $p_0 = p_x$ (hypothèse d'égalité)
 - ◆ Celle que l'on cherche à réfuter
 - ◆ Celle qui est « vraie » tant qu'on a pas démontrer le contraire
- p_x = valeur théorique (population de référence) ou valeur obtenue dans un autre échantillon

Hypothèses à priori

- H1 : hypothèse alternative

- ◆ $p_0 \neq p_x$: bilatéral

- ◆ $p_0 > p_x$: unilatéral

- ◆ $p_0 < p_x$: unilatéral

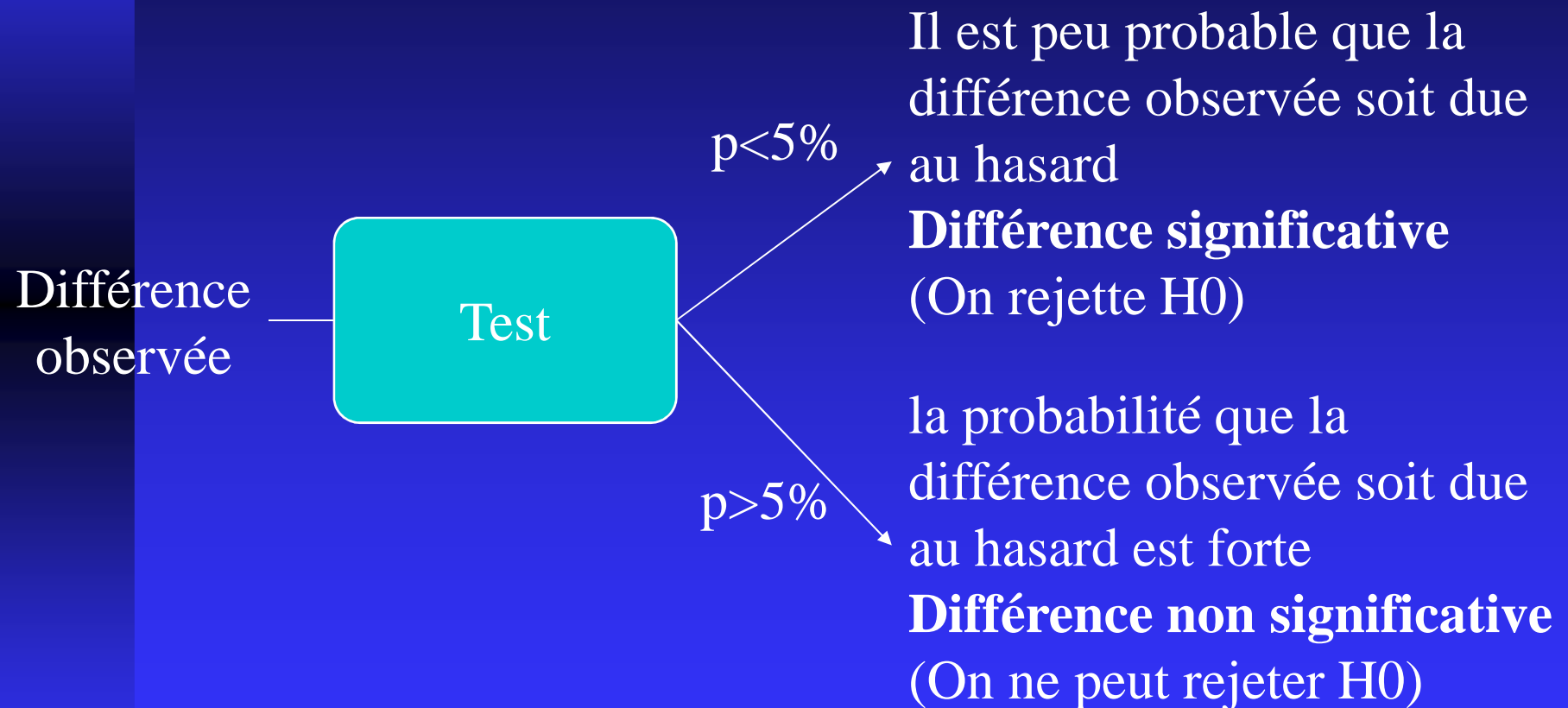
Risques d'erreurs associés à un test

- Risque d'erreur α (1^o espèce) :
 - ◆ probabilité de rejeter H_0 si H_0 est vrai
 - ◆ Probabilité de trouver une différence significative alors qu'il n'y en a pas
- Risque d'erreur β (2^o espèce) :
 - ◆ probabilité de ne pas rejeter H_0 si H_1 est vrai
- Puissance = $1 - \beta$:
 - ◆ **probabilité de rejeter H_0 si H_1 est vrai**
 - ◆ Aptitude à mettre en évidence un différence lorsqu'elle existe

Résultats d'un test

- Valeur du test : la statistique observée
- Comparaison à une valeur théorique donnée par une table
 - ◆ Fonction d'un risque α fixé à priori
- Règle de décision (Neyman-Pearson)
 - ◆ Si statistique observée supérieure à valeur théorique
 - ◆ Test significatif : rejet H_0
 - ◆ Si inférieure
 - ◆ Test non significatif : non rejet de H_0

Interprétation



Degré de signification

p : quantifie le « désaccord » entre ce qu'on observe et H_0

Exemple : comparaison de 2 pourcentages

- % d'insuffisance coronarienne plus élevé chez les diabétiques ?

		Insuf Coro		
		Oui	Non	
Diabète	Oui	39	163	202
	Non	47	679	726
		86	842	928

19,3% d'insuffisance coronarienne chez les diabétiques
6,5% chez les non diabétiques

Est ce que 19,3% est significativement supérieur à 6,5% ?

➤ Test X² à 1 ddl (col-1) * (lignes-1)

Exemple : comparaison de 2 pourcentages

		Insuf Coro			OBSERVE		
		Oui	Non		Oui	Non	
Diabète	Oui	39	163	202	a	b	n3
	Non	47	679	726	c	d	n4
		86	842	928	n1	n2	T
		Insuf Coro			ATTENDU si indépendance		
		Oui	Non		Oui	Non	
Diabète	Oui	18,7	183,3	202	$n1 \cdot n3 / T$	$n2 \cdot n3 / T$	n3
	Non	67,3	658,7	726	$n1 \cdot n4 / T$	$n2 \cdot n4 / T$	n4
		86	842		n1	n2	T

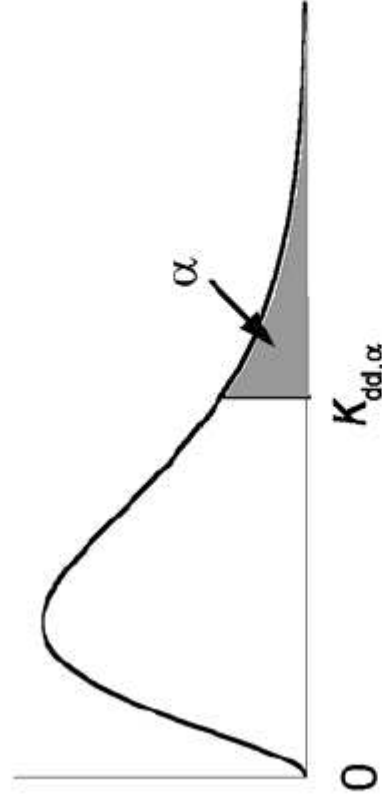
$$\begin{aligned}
 X^2 &= \sum (O_i - E_i)^2 / E_i \\
 &= (18,7 - 39)^2 / 18,7 + (183,3 - 163)^2 / 183,3 + (67,3 - 47)^2 / 67,3 \\
 &\quad + (679 - 658,7)^2 / 679 \\
 &= 30,95
 \end{aligned}$$

La table donne la probabilité α pour que χ^2 égale ou dépasse une valeur donnée, en fonction du nombre de degrés de liberté (d. d. l.)

Quand le nombre de degrés de liberté est élevé,

$\sqrt{2\chi^2}$ est à peu près distribué normalement

autour de $\sqrt{2(\text{d.d.l.}) - 1}$ avec une variance égale à 1



α	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
ddl									
1	0,0158	0,455	1,074	1,642	2,706	3,841	5,412	6,635	10,827
2	0,211	1,386	2,408	3,219	4,605	5,991	7,824	9,210	13,815
3	0,584	2,366	3,665	4,642	6,251	7,815	9,837	11,345	16,266
4	1,064	3,357	4,878	5,989	7,779	9,488	11,668	13,277	18,467
5	1,610	4,351	6,064	7,289	9,236	11,070	13,388	15,086	20,515
6	2,204	5,348	7,231	8,558	10,645	12,592	15,033	16,812	22,457
7	2,833	6,346	8,383	9,803	12,017	14,067	16,622	18,475	24,322
8	3,490	7,344	9,524	11,030	13,362	15,507	18,168	20,090	26,125
9	4,168	8,343	10,656	12,242	14,684	16,919	19,679	21,666	27,877

Seuil décision, degré de signification

■ Seuil de décision

- ◆ $\alpha = 5\%$, 1 ddl
- ◆ Valeur théorique = 3.84
- ◆ La statistique $30.95 > 3.84$: on rejette H_0
- ◆ Différence significative ou les 2 variables ne sont pas indépendants

■ Degré de signification

- ◆ statistique 30.95, 1 ddl
- ◆ $p < 0.001$

Comparaisons entre 2 échantillons (1)

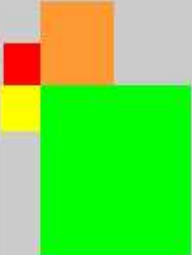
- Comparaison de 2 pourcentages
 - ◆ Variables dichotomiques :
 - ◆ Test X^2 à 1 ddl
 - ◆ Variables à plusieurs classes
 - ◆ Test X^2 à $(l-1)(c-1)$ ddl
 - ◆ Attention si effectifs attendus sous $H_0 < 5$
 - ◆ X^2 corrigé (Yates)
 - ◆ Test Fisher
 - ◆ Si appariement (cas-témoins) : test X^2 appariés

[Next Procedure](#)

[Forward](#)

CORO

diab	1	2	TOTAL
1	39	163	202
Row %	19,3	80,7	100,0
Col %	45,3	19,4	21,8
2	47	679	726
Row %	6,5	93,5	100,0
Col %	54,7	80,6	78,2
TOTAL	86	842	928
Row %	9,3	90,7	100,0
Col %	100,0	100,0	100,0



Lien entre diabète et insuffisance coronarienne

Single Table Analysis

	Point Estimate	95% Confidence Interval	
		Lower	Upper
PARAMETERS: Odds-based			
Odds Ratio (cross product)	3,4566	2,1872	5,4628 (T)
Odds Ratio (MLE)	3,4507	2,1743	5,4596 (M)
		2,1214	5,5924 (F)
PARAMETERS: Risk-based			
Risk Ratio (RR)	2,9823	2,0094	4,4264 (T)
Risk Difference (RD%)	12,8331	7,1031	18,5631 (T)

(T=Taylor series; C=Cornfield; M=Mid-P; F=Fisher Exact)

Previous
 Next
 Last
 History
 Open
 Bookmark
 Print
 Restore

2	47	679	726
Row %	6,5	93,5	100,0
Col %	54,7	80,6	78,2
TOTAL	86	842	928
Row %	9,3	90,7	100,0
Col %	100,0	100,0	100,0

Single Table Analysis

	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
PARAMETERS: Odds-based			
Odds Ratio (cross product)	3,4566	2,1872	5,4628 (T)
Odds Ratio (MLE)	3,4507	2,1743	5,4596 (M)
		2,1214	5,5924 (F)
PARAMETERS: Risk-based			
Risk Ratio (RR)	2,9823	2,0094	4,4264 (T)
Risk Difference (RD%)	12,8331	7,1031	18,5631 (T)

(T=Taylor series; C=Cornfield; M=Mid-P; F=Fisher Exact)

STATISTICAL TESTS	Chi-square	1-tailed p	2-tailed p
Chi square - uncorrected	30,9520		0,0000000000
Chi square - Mantel-Haenszel	30,9186		0,0000000000
Chi square - corrected (Yates)	29,4446		0,0000012268
Mid-p exact		0,0000001480	
Fisher exact		0,0000002320	

Comparaisons entre 2 échantillons (2)

- Une variable quantitative + une variable qualitative
 - ◆ Comparaison de 2 moyennes
 - ◆ Test t de Student à n_1+n_2-2 ddl
 - ◆ Comparaison de plusieurs moyennes
 - ◆ ANOVA (test F)
 - ◆ Attention si variances différentes
 - ◆ Si appariement (cas-témoins) : test Student à $n-1$ ddl

Poids selon Statut diabétique

Descriptive Statistics for Each Value of Crosstab Variable

	Obs	Total	Mean	Variance	Std Dev	
no	742	44972,2000	60,6094	172,7525	13,1435	
yes	205	13830,5000	67,4659	171,4670	13,0945	
	Minimum	25%	Median	75%	Maximum	Mode
no	17,0000	52,0000	60,0000	69,0000	130,0000	67,0000
yes	35,0000	59,0000	66,0000	74,5000	121,0000	65,0000

ANOVA, a Parametric Test for Inequality of Population Means

(For normally distributed data only)

Variation	SS	df	MS	F statistic
Between	7550,9670	1	7550,9670	43,7801
Within	162988,8750	945	172,4750	
Total	170539,8420	946		

T Statistic = 6,6167

P-value = 0,0000

Bartlett's Test for Inequality of Population Variances

Bartlett's chi square= 0,0044 df=1 P value=0,9468

Comparaisons entre 2 échantillons (3)

- Deux variables quantitatives
 - ◆ Coefficient de corrélation linéaire
 - ◆ test Student à $n - 2$ ddl
 - ◆ Attention si relation non linéaire

[Previous Dataset Results Library](#)

REGRESS Poids = Tail

Lien entre poids et taille

[Next Procedure](#)

Linear Regression

Variable	Coefficient	Std Error	F-test	P-Value
Tail	0,712	0,038	344,5371	0,000000
CONSTANT	-54,911	6,309	75,7552	0,000000

Correlation Coefficient: $r^2 = 0,28$

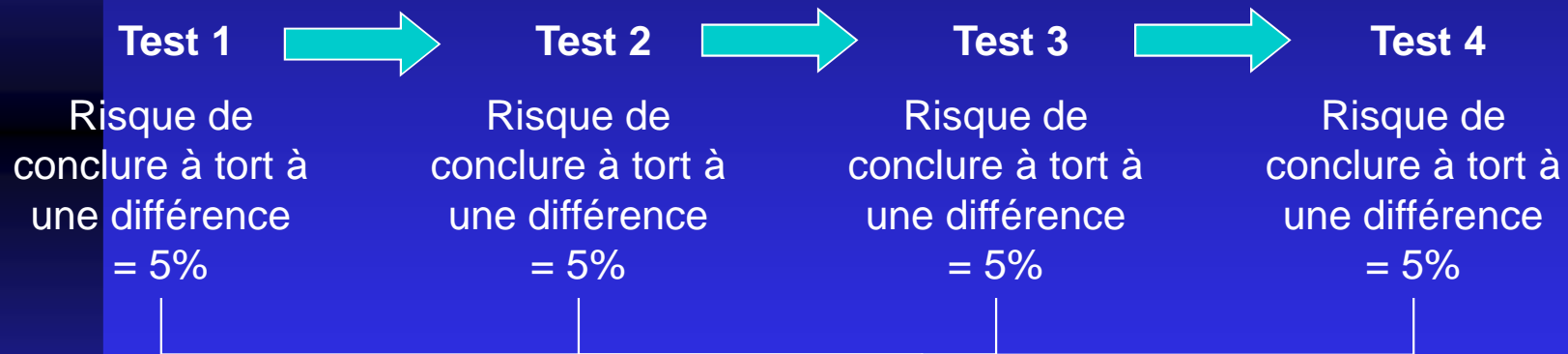
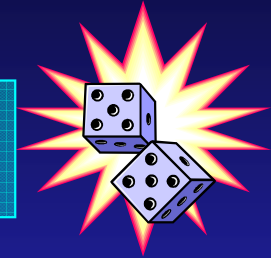
Source	df	Sum of Squares	Mean Square	F-statistic
Regression	1	47858,171	47858,171	344,537
Residuals	900	125015,154	138,906	
Total	901	172873,325		

Application à l'essai thérapeutique

- L'essai clinique est utilisé pour montrer l'effet d'un traitement
- **Risque alpha** : considérer comme efficace un traitement qui ne l'est pas
- **Risque bêta** : ne pas conclure alors que le traitement est efficace
- **Puissance** : montrer l'efficacité d'un traitement réellement efficace

Comparaisons multiples

Aux dés, la probabilité d'obtenir un six est plus forte avec 3 dés qu'avec un seul



n	risque global
2	0.10
3	0.13
5	0.23
10	0.40

Globalement, le risque de conclure à tort à une différence lors de ces 4 comparaisons est bien plus important que 5%.

Inflation du risque alpha

Répétition des tests

- Conclusion basée
 - ◆ non pas sur un seul test
 - ◆ mais sur plusieurs
- Conclusion à un effet à partir du moment où il existe au moins un test significatif
- Le risque α de la conclusion est bien supérieure à 5%
- Inflation du risque alpha

Critère de jugement principal

- Décès de toute cause
- Décès cardiovasculaire
- Mort subite
- Infarctus
- Accident vasculaire cérébraux
- Chirurgie



Pas de définition
de critère principal

7 tests statistiques

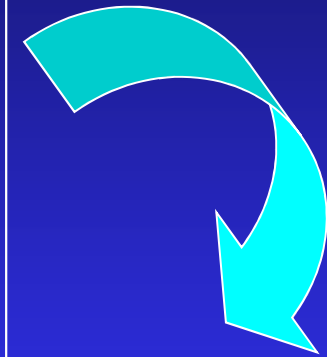
Risque de conclure à
tort à l'efficacité du
traitement = 30%

Critère principal

- Décès de toute cause

Critères secondaires

- Décès cardiovasculaire
- Mort subite
- Infarctus
- Accident vasculaire cérébraux
- Chirurgie



Définition a priori
d'un critère principal

Un seul test statistique

Risque de conclure à
tort à l'efficacité du
traitement = 5%

Essai de confirmation

- Formulation a priori d'une hypothèse
 - ◆ objectif de l'essai
- Démarche hypothético-déductive
 - ◆ confrontation à la réalité d'une hypothèse par une expérience conçue spécialement pour cela
 - ◆ Valeur méthodologique du résultat

Essai exploratoire

- Pas d'hypothèse préalablement fixé
- Grand nombre de comparaison possible
 - ◆ inflation du risque alpha
- Epistémologiquement non valide
 - ◆ induction
 - ◆ observation d'un fait particulier --> loi générale
- Pêche à la ligne
- Sert à générer de nouvelles hypothèses

Hypothèse post hoc

- Formulée d'après les résultats
- Vérifiée sur les mêmes données
- Situation tautologique

Analyses en sous groupes

- De nature exploratoire
- Ne donne pas de démonstration
- Suggèrent des variations d'efficacité

Différence significative

Résultat
significatif



Significativité
clinique

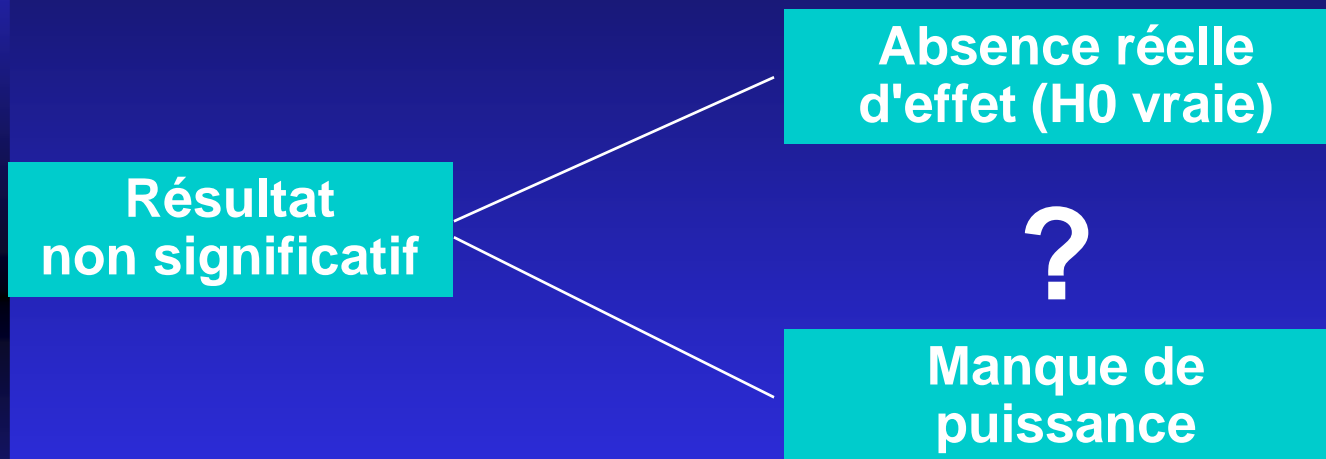
?

- Taille d'effet
- Degré de signification (p) : ne renseigne pas sur l'importance de la différence !!

Validité interne

- le résultat est très probablement réel
 - ◆ il est statistiquement significatif
 - ◆ obtenu sur le critère de jugement principal
 - ◆ il est méthodologiquement valide (hypothèse formulée a priori)
 - ◆ il n'a pas été obtenu par une analyse en sous groupes
- le résultat est sûr (exempt de biais)
 - ◆ le plan d'expérience choisi évite les biais
 - ◆ et l'étude a été correctement réalisée

Différence non significative



- Impossible de conclure
- Ne prouve pas qu'il n'y a pas d'effet

Puissance

- La seule chose que l'on démontre par des observations est qu'une hypothèse est fautive. On ne peut démontrer qu'elle est vraie. « Elle n'est pas contredite par l'expérience ».
- Minimisation de α : évidence forte en faveur de l'hypothèse nulle
- Mais aussi maximiser la puissance

Calculer la puissance d'une étude à priori

- Approximation car repose sur des estimations : tests bruts, écarts attendus, suivis, données manquantes...
- Calcul de la puissance
- Calcul du nombre de sujets nécessaires
- Calcul de la différence minimum que l'on peut mettre en évidence
 - Éviter une puissance à priori $< 80\%$
 - càd prendre un risque $\beta < 20\%$

Comment améliorer la puissance d'une étude

- Si α diminue, puissance diminue
- Unilatéral > bilatéral (puissance augmente)
- Si grande variance, puissance faible : homogénéité des sujets, stratification, ajustement (attention surajustement !)
- Puissance maximale quand les effectifs des 2 groupes sont égaux (cohorte, cas-témoins)
- Si différence à mettre en évidence est grande, puissance plus grande : contraste des groupes comparés (attention biais !)
- Puissance d'autant plus grande que le nombre de sujets est important : attention PDV, données manquantes
- Biais de mesure : dilution

Test unilatéral / bilatéral

- Si unilatéral : augmentation de la puissance, moins de sujets nécessaires
- Mais !!!!
- Si H_1 $p_1 > p_2$ et que l'on observe $p_1 < p_2$
 - ◆ Non rejet de H_0 ($P_1 = P_0$) quelque soit l'écart entre p_1 et p_2

Exemple : Lupus

- Critère de jugement principal : nombre de rémissions à 6 mois
- TT1 : faibles doses , TT2 : fortes doses
- Objectif : montrer qu'avec TT1 ont a moins de rémissions qu'avec TT2
 - ◆ Comparaison de 2 proportions
 - ◆ $H_0 : P_1 = P_2$
 - ◆ $H_1 : P_1 < P_2$ (test unilatéral)
 - ◆ $H_1 : P_1 \neq P_2$ (test bilatéral)

Calcul du nombre de sujets nécessaires pour la comparaison de 2 proportions

$$N_1 = \frac{k+1}{k} \frac{(z_{\alpha/2} - z_{1-\beta})^2}{2 (\arcsin \sqrt{P_2} - \arcsin \sqrt{P_1})^2}$$

Table de la loi normale centrée réduite test bilatéral

$$N_2 = k N_1$$

% vraie sous H_0 et H_1

Si effectifs inégaux

Calcul du nombre de sujets nécessaires (unilatéral)

$$N_1=N_2= \frac{(z_{\alpha} - z_{1-\beta})^2}{2 (\arcsin \sqrt{P_2} - \arcsin \sqrt{P_1})^2}$$

- Effectifs égaux dans chaque groupe
- Puissance minimale 80%

$$(z_{\alpha} - z_{1-\beta})^2 = (1,645 + 0,842)^2 = (2,487)^2 = 6,185$$

$$P_1 = 50\%, \arcsin \sqrt{P_1} = 0,785$$

$$P_2 = 70\%, \arcsin \sqrt{P_2} = 0,991$$

$$N_1=N_2= 6,185 / 0,084872 = 73$$

Calcul du nombre de sujets nécessaires (bilatéral)

$$N_1=N_2= \frac{(z_{\alpha/2} - z_{1-\beta})^2}{2 (\arcsin \sqrt{P_2} - \arcsin \sqrt{P_1})^2}$$

- Effectifs égaux dans chaque groupe
- Puissance minimale 80%

$$(z_{\alpha/2} - z_{1-\beta})^2 = (1,960 + 0,842)^2 = (2,802)^2 = 7,851$$

$$P_1 = 50\%, \arcsin \sqrt{P_1} = 0,785$$

$$P_2 = 70\%, \arcsin \sqrt{P_2} = 0,991$$

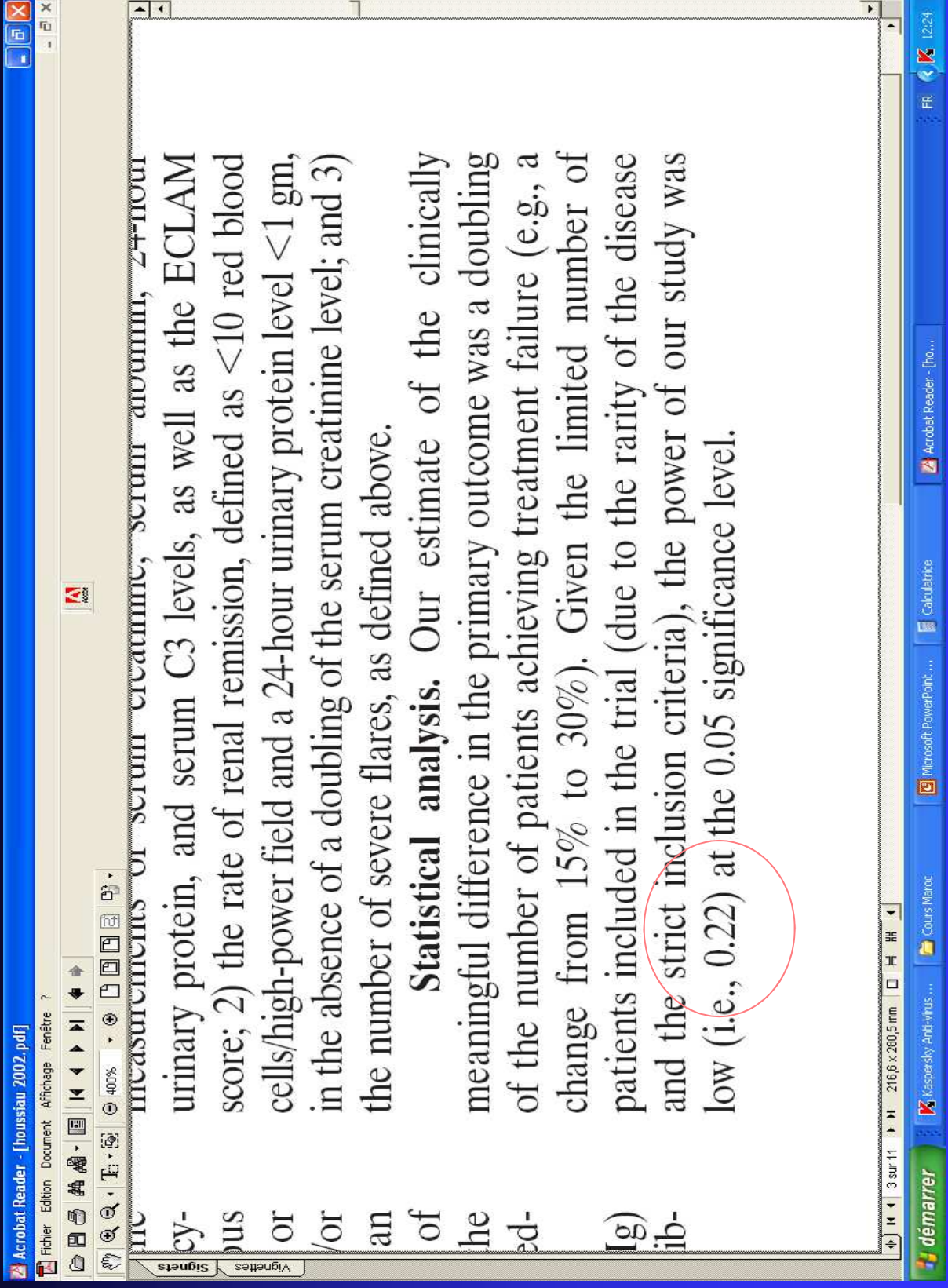
$$N_1=N_2= 7,851 / 0,084872 = 93$$

Houssiau

- Critère de jugement : échec du traitement
- A priori : $P_1 = 30\%$, $P_2 = 15\%$
 - ◆ Si puissance à priori 80%
 - ◆ Bilatéral : 119 patients dans chaque groupe
 - ◆ Unilatéral : 94 patients dans chaque groupe
 - ◆ Avec 45 patients dans chaque groupe
 - ◆ $|\arcsin \sqrt{P_2} - \arcsin \sqrt{P_1}| / \sqrt{1/2n} = 1,72$
 - ◆ Puissance à priori = 40% bilatéral, 53% unilatéral
- A postérieur : $P_1 = 16\%$, $P_2 = 20\%$
 - ◆ Avec 45 patients dans chaque groupe
 - ◆ Puissance à postérieur = 3% bilatéral, 5% unilatéral!!!!
- Au final, que 40 et 38 patients dans les groupes.....

Conséquences

- Manque de puissance : ne pas rejeter H_0 si H_1 est vrai
 - ◆ Houssiau : La probabilité que le test de comparaison des fréquences soit significatif est égale à 22%
 - ◆ Il y avait 22% de chance que cette étude puisse monter un effet !!!
 - ◆ Il y avait 78% de risque que l'on conclut à l'absence de différence



measurements of serum creatinine, serum albumin, 24-hour urinary protein, and serum C3 levels, as well as the ECLAM score; 2) the rate of renal remission, defined as <10 red blood cells/high-power field and a 24-hour urinary protein level <1 gm, in the absence of a doubling of the serum creatinine level; and 3) the number of severe flares, as defined above.

Statistical analysis. Our estimate of the clinically meaningful difference in the primary outcome was a doubling of the number of patients achieving treatment failure (e.g., a change from 15% to 30%). Given the limited number of patients included in the trial (due to the rarity of the disease and the strict inclusion criteria), the power of our study was low (i.e., 0.22) at the 0.05 significance level.

Bibliographie

Falissard B.

Comprendre et utiliser les statistiques dans les sciences de la vie.
2ème édition.

Paris : Masson; 1998. 332 pages (Abrégés).

Bouyer J, Hémon D, Cordier S, Derriennic F, Stücker I, Stengel B,
Clavel J.

Epidémiologie. Principes et méthodes quantitatives.

Paris : INSERM; 1995. 498 pages.

Schwartz D.

Méthodes statistiques à l'usage des médecins et des biologistes.
4ème édition. Flammarion 1995.

Bouyer J.

Méthodes statistiques. Médecine - Biologie.

Paris : INSERM; 1996. 353 pages.

Tests paramétriques en fonction du type des variables et des effectifs

		Variable	
		Qualitative	Quantitative
Variable qualitative	Echantillons ou groupes de patients indépendants	<u>Chi2</u> ou z n=2	<u>t de Student</u> ou z Analyse de variance à un facteur
	Echantillons ou groupes de patients appariés	<u>Chi2</u> n>2	Analyse de variance à deux facteurs <u>t de Student</u>
		<u>Chi2 de McNemar</u> n=2	Analyse de variance
		Stuart-Maxwell ou <u>Chi2 de Mantel-Haenszel</u> n>2	Coefficient de corrélation de <u>Pearson</u>
Variable quantitative		-	-

n : nombre d'échantillons

Tests non paramétriques en fonction du type des variables et des effectifs

		Variable	
		Qualitative	Quantitative
Variable qualitative	Echantillons ou groupes de patients indépendants	n=2 Test exact de Fisher	Wilcoxon ou Mann-Whitney
		n>2 Test exact de Fisher	Kruskal-Wallis
	Echantillons ou groupes de patients appariés	n=2 Chi2 de McNemar	Wilcoxon signé Friedman
		n>2 Cochran Q ou Chi2 de Mantel-Haenszel	Test de Friedman
Variable quantitative		-	Coefficient de corrélation des rangs de Spearman ou de Kendall