BIOSTATISTIQUE

Ch. Mélot, MD, PhD, MSciBiostat
Faculté de Médecine
Université Libre de Bruxelles
et
Service des Soins Intensifs
Hôpital Universitaire Erasme
Bruxelles, Belgique

cmelot@ulb.ac.be

DEA cardiopneumo, Paris le 12 février 2007

Un statisticien est une personne qui peut avoir la tête dans un four et les pieds pris dans la glace et dire qu'en « moyenne » il se sent bien.

Benjamin DERECA

There are three kinds of lies:

Lies

Damned lies

And statistics

B. Disraeli (1804 – 1881)

C.R. RAO

Statistics and Truth. Putting Chance to Work, 2nd Edition, 1997

Uncertain knowledge

+

Knowledge of the amount of uncertainty in it

=

Usable knowledge

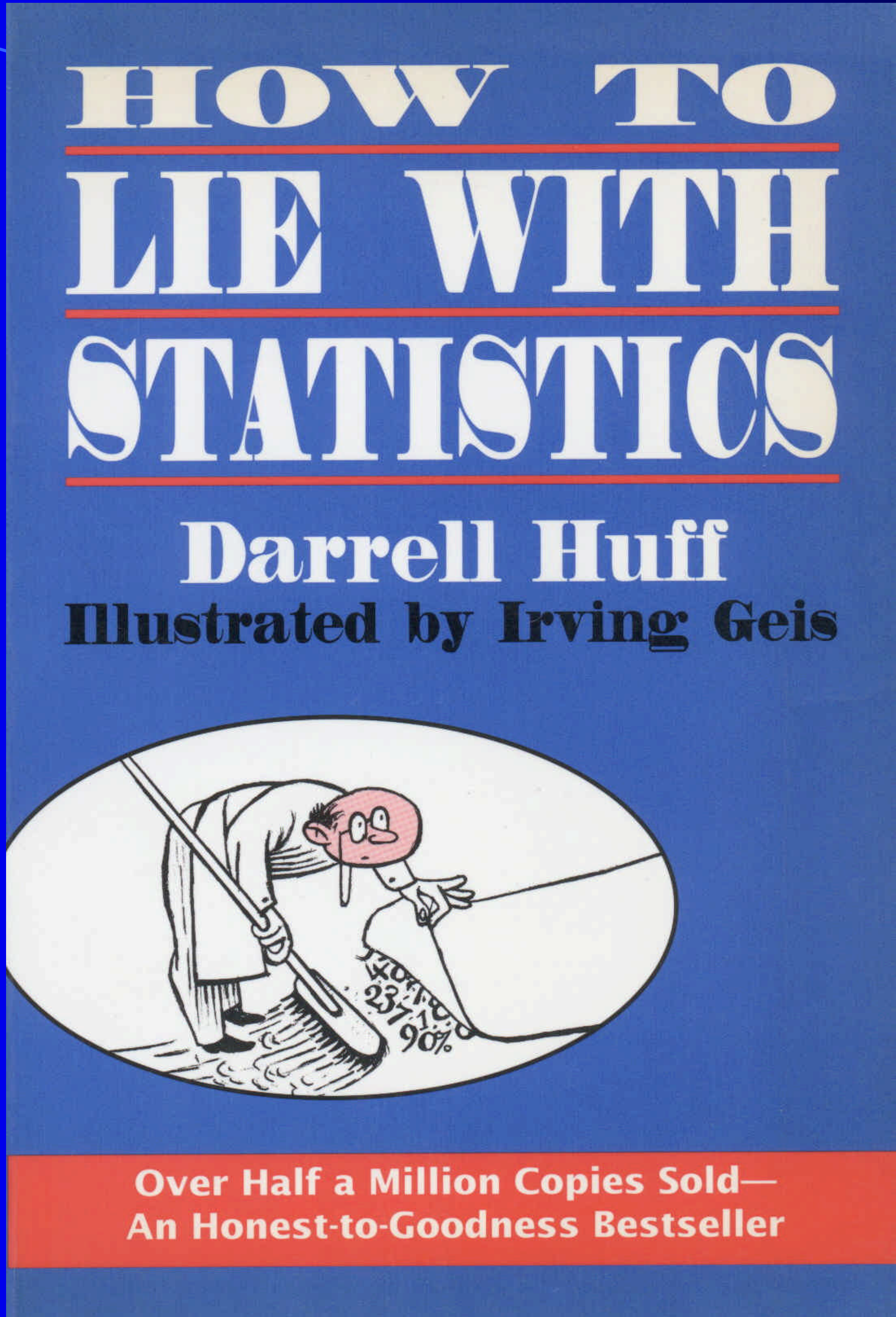
If it were not for the great variability among individuals, medicine might as well be a science and not an art"

Sir William Osler

If I had only one day to live

I would spend it in a statistics class

…that way it would seem longer


HOW TO
LIE WITH
STATISTICS
Darrell Huff
Illustrated by Irving Geis
Over Half a Million Copies Sold—
An Honest-to-Goodness Bestseller

NOTIONS DE BASE EN STATISTIQUE

- HASARD: dérivé de l'arabe "az-zhar", qui signifie "dé à jouer".
  Guillaume de Tyr (1130 - 1184): lors du siège du château de El Azar (Syrie), jeu de dés "Azar" inventé par les croisés.

- ALEATOIRE: "alea" signifie "dé à jouer" en latin

- RANDOMISATION: "at random" signifie "au hasard" en anglais

- STOCHASTIQUE (στοχαστής = devin)

Qu'appelle-t-on nombres aléatoires?

- Considérons le nombre 5:
  - Il est raisonnable de le considérer comme aléatoire s'il constitue le résultat d'un jet de dé, mais pas lorsqu'il apparaît dans la suite des nombres entiers positifs rangés dans l'ordre: 1, 2, 3, 4, 5, 6, 7, 8, 9, … $\rightarrow n_{i+1} = n_i + 1$

- Par essence, une suite de nombres est aléatoire lorsqu'il est impossible de la caractériser de manière plus courte qu'en lisant tous ses éléments.

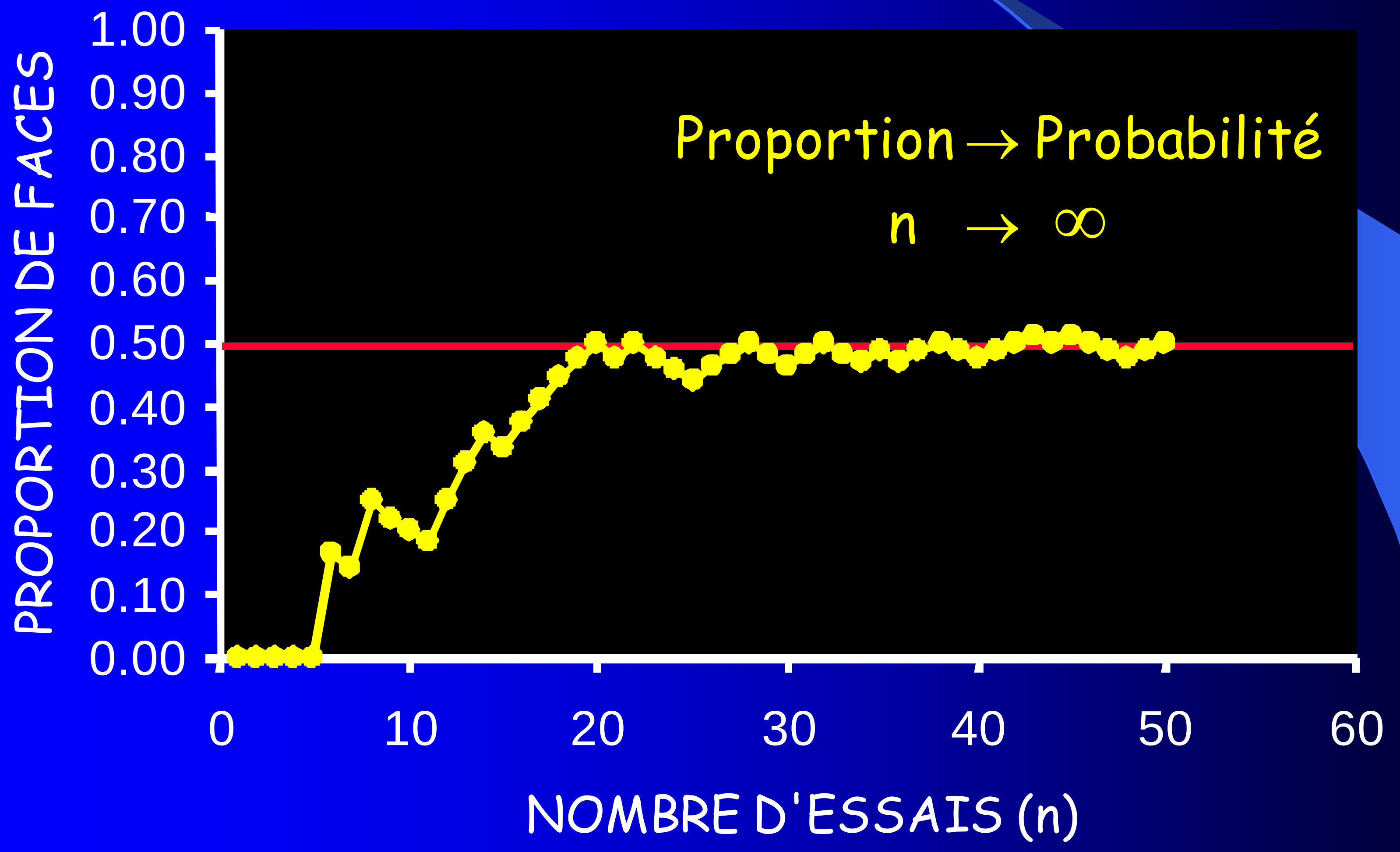
- Méthodes pour engendrer des nombres aléatoires: pièce de monnaie, dé, jeu de carte, roulette, générateur informatique.

Probabilité

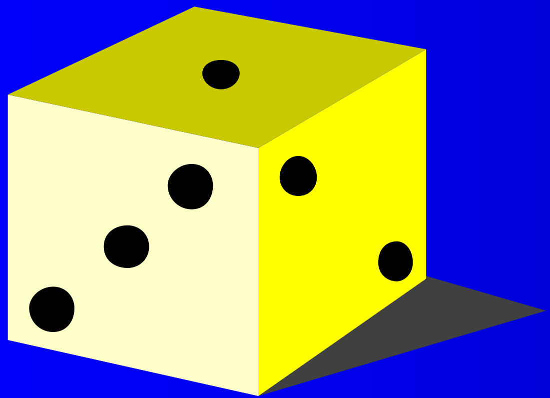
- La probabilité est un nombre entre 0 (0 % de chance) et 1 (100 % de chance)
- Probabilité subjective
- Probabilité statistique (approche fréquentiste)

Probabilité statistique

JET D'UNE PIECE



Odds (cote) et Probabilité



$$\text{Probabilité} = \frac{1}{6} = 0.166$$

$$\text{Odds en faveur} = \frac{\frac{1}{6}}{\frac{5}{6}} = 0.20$$

$$\text{Odds contre} = \frac{\frac{5}{6}}{\frac{1}{6}} = 5 \text{ contre } 1$$

TYPES DE VARIABLES (1)

- Variables qualitatives catégorielles:
  - nominales
    - par exemple:
      - caractère génétique: wild type / knock-out
      - sexe: masculin / féminin
  - ordinales
    - par exemple:
      - groupes d 'âge: 10-19; 20-29;...
      - groupes traités par doses croissantes d 'un médicament

TYPES DE VARIABLES (2)

- Variables quantitatives:
  - discrètes (dénombrements)
    - par exemple:
      - nombre de patients décédés
      - nombre de récidives
  - continues (mesures ou proportions)
    - par exemple:
      - poids (kg) , taille (cm) , âge (semaines), pression artérielle systolique (mmHg)
      - proportion de cellules apoptotiques (variable continue entre 0 et 1, sans dimension)

CODAGE DES VARIABLES (3)

- Variables qualitatives:
  - Doivent être transformées en valeur numérique:
    - exemple:
      - Fumeur: Oui = 1; Non = 0
      - Sexe_M : Féminin = 0; Masculin = 1
      - Groupes d'âge: 10-19 = 1; 20-29 = 2;...
- Variables quantitatives:
  - Continues (mesures, proportions)
    - exemple:
      - Poids (kg): 55; 65;...
      - proportion de décès: 0.12; 0.14
  - Discrètes ou catégorielles (comptages)
    - Exemple:
      - fumeurs: 10 ; non fumeurs: 90

CODAGE DES VARIABLES (4)

Antécédents	~~ATCD~~	ATC_Infa	ATC_Asth	ATC_HTA	ATC_BPCO
Infarctus	~~1~~	1	0	0	0
Asthme	~~2~~	0	1	0	0
HTA	~~3~~	0	0	1	0
BPCO	~~4~~	0	0	0	1

Création d'une base de données (1)

PatID	AGE	SEXE_M	TABAC	ATD_HTA	ATD_BPC	ATD_INFA	V1_CREA	V1_Na	V2_CREA	V2_Na	V2_MOIS
1	32	1	0	0	1	0	1.2	134	0.9	135	2
2	45	0	1	1	0	0	1.3	128	0.8	135	3
3	56	0	0	0	0	0	0.8	134	0.5	132	2
4	76	0	2	1	0	0	1.2	135	1.1	132	3
5	34	1	2	0	1	0		143	1.2	131	4
6	86	1	1	0	0	1	2.1	136	1.4	138	2.5
7	34	0	1	0	0	0	1.1	135	1.5	140	3.5
8	67	1	2	0	0	0	0.7	135	1.6	141	2.7
9	98	0	1	0	0	0	0.9	132	1.2	142	2.8
10	54	0	2	1	0	0	0.8	132	1.3	145	3
11	65	1	0	0	1	1	0.5	131	0.8	134	4
12	67	1	0	0	0	0	1.1	138	1.2	128	3
13	68	1	1	0	1	0	1.2	140	1.3	134	5
14	65	0	0	0	0	1	1.4	141	2.1	135	3
15	46	1	2	1	0	0	1.5	142	1.1	143	4

Nom des variables:
- 1ère ligne
- maximum 8 caractères, sans espace
- chiffres, lettres (non accentuées)
- certains caractères spéciaux: _
- nom différent pour chaque variable

Création d'une base de données (1)

Microsoft Excel - DBexemple.xls

K2 fx 135

	PatID	AGE	SEXE_M	TABAC	ATD_HTA	ATD_BPC	ATD_INFA	V1_CREA	V1_Na	V2_CREA	V2_Na	V2_MOIS
2	1	32	1	0	0	1	0	1.2	134	0.9	135	2
3	2	45	0	1	1	0	0	1.3	128	0.8	135	3
4	3	56	0	0	0	0	0	0.8	134	0.5	132	2
5	4	76	0	2	1	0	0	1.2	135	1.1	132	3
6	5	34	1	2	0	1	0		143	1.2	131	4
7	6	86	1	1	0	0	1	2.1	136	1.4	138	2.5
8	7	34	0	1	0	0	0	1.1	135	1.5	140	3.5
9	8	67	1	2	0	0	0	0.7	135	1.6	141	2.7
10	9	98	0	1	0	0	0	0.9	132	1.2	142	2.8
11	10	54	0	2	1	0	0	0.8	132	1.3	145	3
12	11	65	1	0	0	1	1	0.5	131	0.8	134	4
13	12	67	1	0	0	0	0	1.1	138	1.2	128	3
14	13	68	1	1	0	1	0	1.2	140	1.3	134	5
15	14	65	0	0	0	0	1	1.4	141	2.1	135	3
16	15	46	1	2	1	0	0	1.5	142	1.1	143	4

Données:
- 1 rangée = 1 unité expérimentale (patient, animal, …)
- uniquement des chiffres
- si < limite de détection: < 0.02 → choisir 0.01 p.ex
- séparateur décimal: 10.0 (point)
- si donnée manquante: laisser la cellule vide

Création d'une base de données (2)

Statistix - [Untitled]

File Edit Data Statistics Preferences Window Help

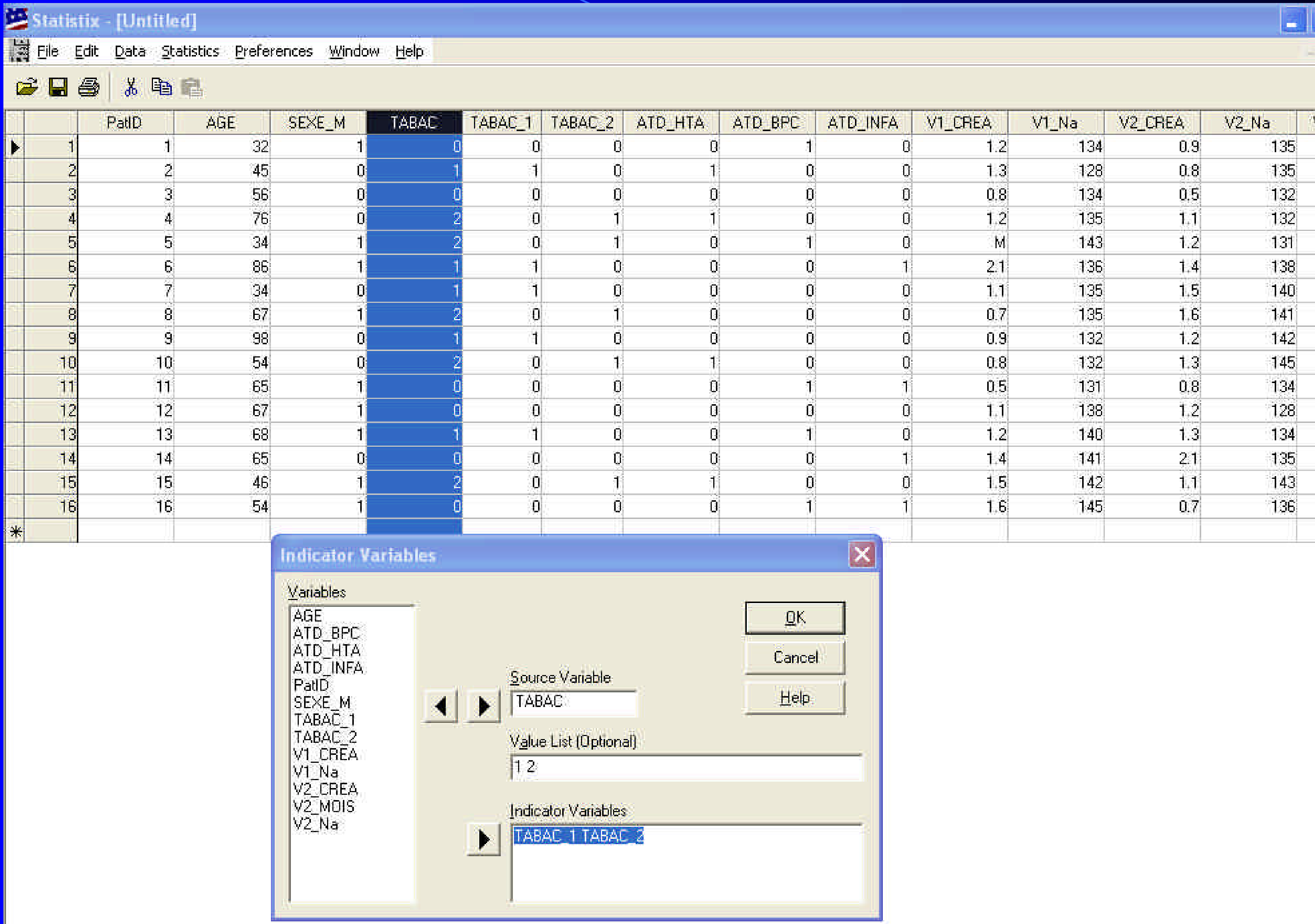
	PatID	AGE	SEXE_M	TABAC	ATD_HTA	ATD_BPC	ATD_INFA	V1_CREA	V1_Na	V2_CREA	V2_Na	V2_MOIS
1	1	32	1	0	0	1	0	1.2	134	0.9	135	2
2	2	45	0	1	1	0	0	1.3	128	0.8	135	3
3	3	56	0	0	0	0	0	0.8	134	0.5	132	2
4	4	76	0	2	1	0	0	1.2	135	1.1	132	3
5	5	34	1	2	0	1	0	M	143	1.2	131	4
6	6	86	1	1	0	0	1	2.1	136	1.4	138	2.5
7	7	34	0	1	0	0	0	1.1	135	1.5	140	3.5
8	8	67	1	2	0	0	0	0.7	135	1.6	141	2.7
9	9	98	0	1	0	0	0	0.9	132	1.2	142	2.8
10	10	54	0	2	1	0	0	0.8	132	1.3	145	3
11	11	65	1	0	0	1	1	0.5	131	0.8	134	4
12	12	67	1	0	0	0	0	1.1	138	1.2	128	3
13	13	68	1	1	0	1	0	1.2	140	1.3	134	5
14	14	65	0	0	0	0	1	1.4	141	2.1	135	3
15	15	46	1	2	1	0	0	1.5	142	1.1	143	4
16	16	54	1	0	0	1	1	1.6	145	0.7	136	3

CREATION DE VARIABLES FICTIVES (DUMMY VARIABLES)

	Tabac	Tabac_1	Tabac_2
Non fumeur	0	0	0
Fumeur (≤ 1 paquet/j)	1	1	0
Gros fumeur (> 1 paquet/j)	2	0	1

Nombre de niveaux - 1 = Nombre de variables fictives

Création d'une base de données (3)



	PatID	AGE	SEXE_M	TABAC	TABAC_1	TABAC_2	ATD_HTA	ATD_BPC	ATD_INFA	V1_CREA	V1_Na	V2_CREA	V2_Na
1	1	32	1	0	0	0	0	1	0	1.2	134	0.9	135
2	2	45	0	1	1	0	1	0	0	1.3	128	0.8	135
3	3	56	0	0	0	0	0	0	0	0.8	134	0.5	132
4	4	76	0	2	0	1	1	0	0	1.2	135	1.1	132
5	5	34	1	2	0	1	0	1	0	M	143	1.2	131
6	6	86	1	1	1	0	0	0	1	2.1	136	1.4	138
7	7	34	0	1	1	0	0	0	0	1.1	135	1.5	140
8	8	67	1	2	0	1	0	0	0	0.7	135	1.6	141
9	9	98	0	1	1	0	0	0	0	0.9	132	1.2	142
10	10	54	0	2	0	1	1	0	0	0.8	132	1.3	145
11	11	65	1	0	0	0	0	1	1	0.5	131	0.8	134
12	12	67	1	0	0	0	0	0	0	1.1	138	1.2	128
13	13	68	1	1	1	0	0	1	0	1.2	140	1.3	134
14	14	65	0	0	0	0	0	0	1	1.4	141	2.1	135
15	15	46	1	2	0	1	1	0	0	1.5	142	1.1	143
16	16	54	1	0	0	0	0	1	1	1.6	145	0.7	136

TERMINOLOGIE

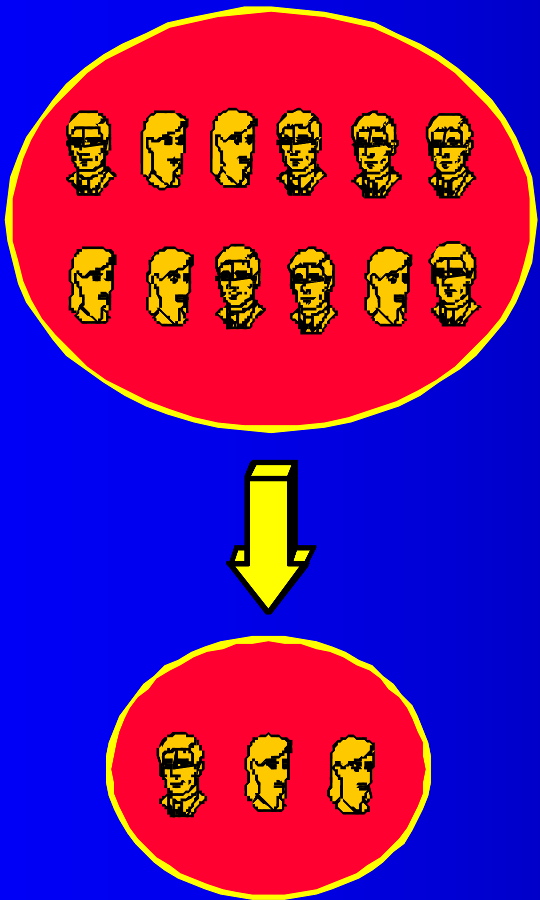
- VARIABLE: Ce terme est utilisé pour caractériser tout ce qui varie dans un ensemble de données.

  Une variable prendra pour un sujet - ou pour un temps donné - une certaine valeur.

  Une variable sera dite aléatoire si les valeurs qu 'elle peut prendre fluctuent au hasard.

- PARAMETRE: Les paramètres permettent de résumer les valeurs prises par une variable:
  - paramètre de tendance centrale (Moyenne)
  - paramètre de dispersion (Ecart-type, «Standard Deviation»)

POPULATION et ECHANTILLON



population (μ, σ)

(une collection complète d'individus, d'objets ou de mesures)

Echantillon (m, s)

(une partie ou un sous-ensemble de la population)

Le but de l'analyse statistique est d'évaluer une population à partir d'échantillons tirés de cette population

- Le résumé des données: la réduction aux paramètres

RESUME DES DONNEES: REDUCTION AUX PARAMETRES.

Paramètre de tendance centrale (1er paramètre): la moyenne arithmétique

	SGOT, UI/L		SGPT,UI/L
	12		52
	15		45
	17		42
	22		39
	23		25
Somme	89		203
n	5		5
Moyenne	17.8		40.6

RESUME DES DONNEES: REDUCTION AUX PARAMETRES.

Paramètre de dispersion (2ème paramètre) : la variance (SD^2) et l'écart-type ('standard deviation') (SD):

	X = SGOT		Y = SGPT	$(X_i - moy)$	$(Y_i - moy)$	$(X_i - moy)^2$	$(Y_i - moy)^2$	
	12		52	-5.8	11.4	33.6	130.0	
	15		45	-2.8	4.4	7.8	19.4	
	17		42	-0.8	1.4	0.6	2.0	
	22		39	4.2	-1.6	17.6	2.6	
	23		25	5.2	-15.6	27.0	243.4	
Somme	89		203	0	0	86.8	397.2	SCE
n	5		5			4	4	ddl
Moyenne	17.8		40.6			21.7	99.3	SD^2
						4.6	9.9	SD

SCE = somme des carrés des écarts (SS = Sum of Squares)
ddl = degré de liberté (df = degree of freedom)

STRUCTURE DE LA VARIANCE

Somme des carrés des écarts (SCE)
(Sum of squares (SS))

Carré moyen
(Mean square (MS))

$$Var(x) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n - 1}$$

Degré de liberté (ddl)
Degree of freedom (df)

DEGRE DE LIBERTE

1er paramètre: ddl = n

$$\text{Moyenne}(x) = \frac{\sum_{i=1}^{n} X_i}{n}$$

Degré de liberté

2ème paramètre: ddl = n-1

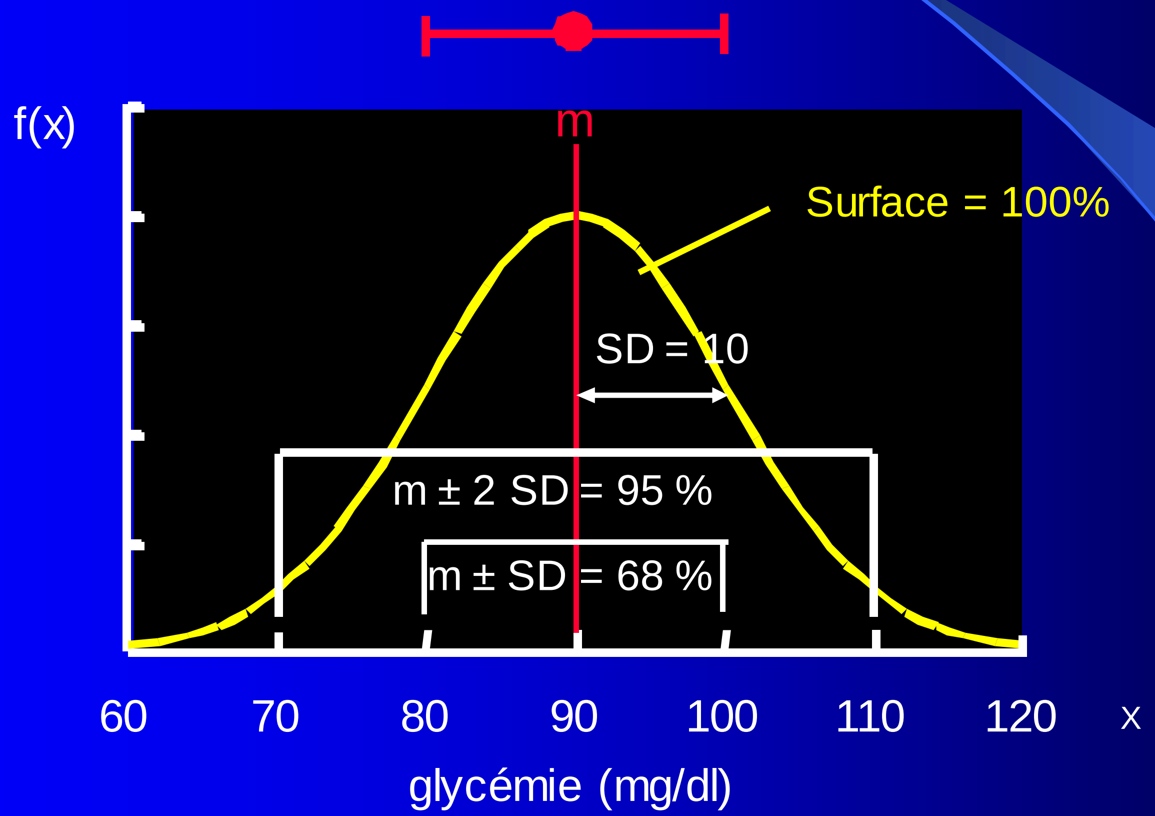
$$\text{Var}(x) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n - 1}$$

Degré de liberté

- La connaissance de la structure des données est importante car elle peut changer l 'interprétation des résultats

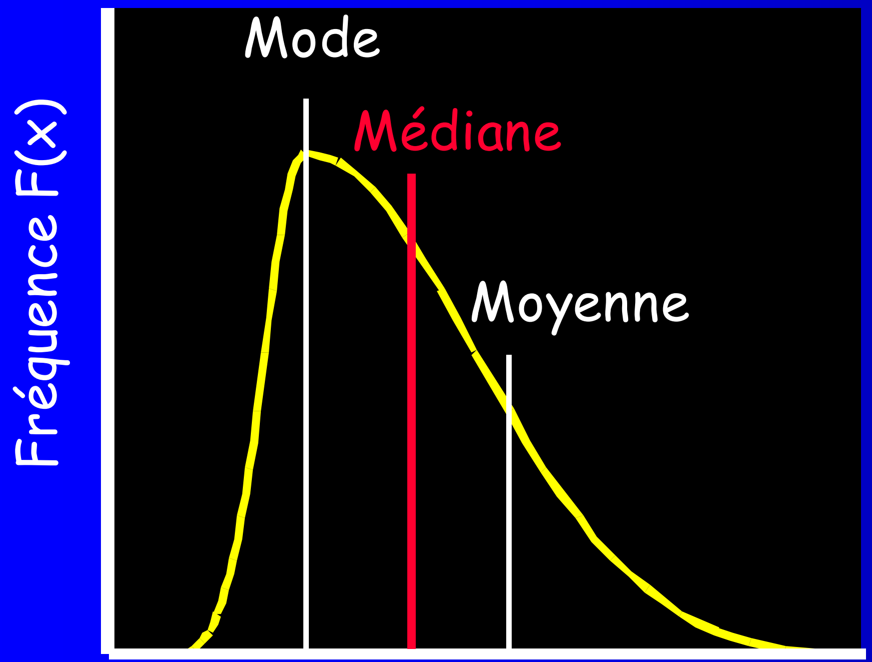
Distribution normale (gaussienne)

Glycémie: 90 ± 10 (moyenne ± SD)

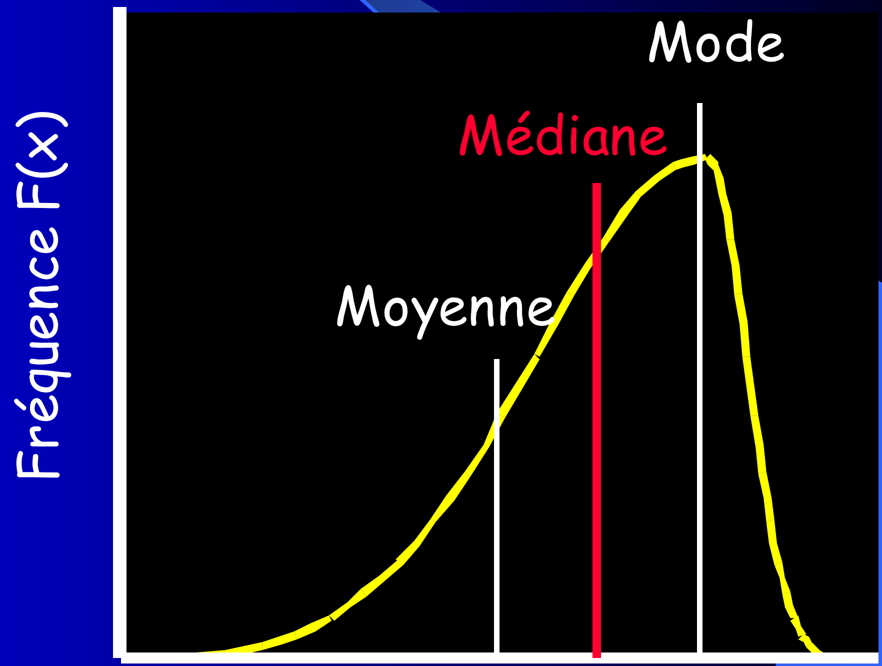


Valeurs normales pour la glycémie: 70 à 110 mg/dl
Intervalle de prédiction à 95 %

Distributions asymétriques

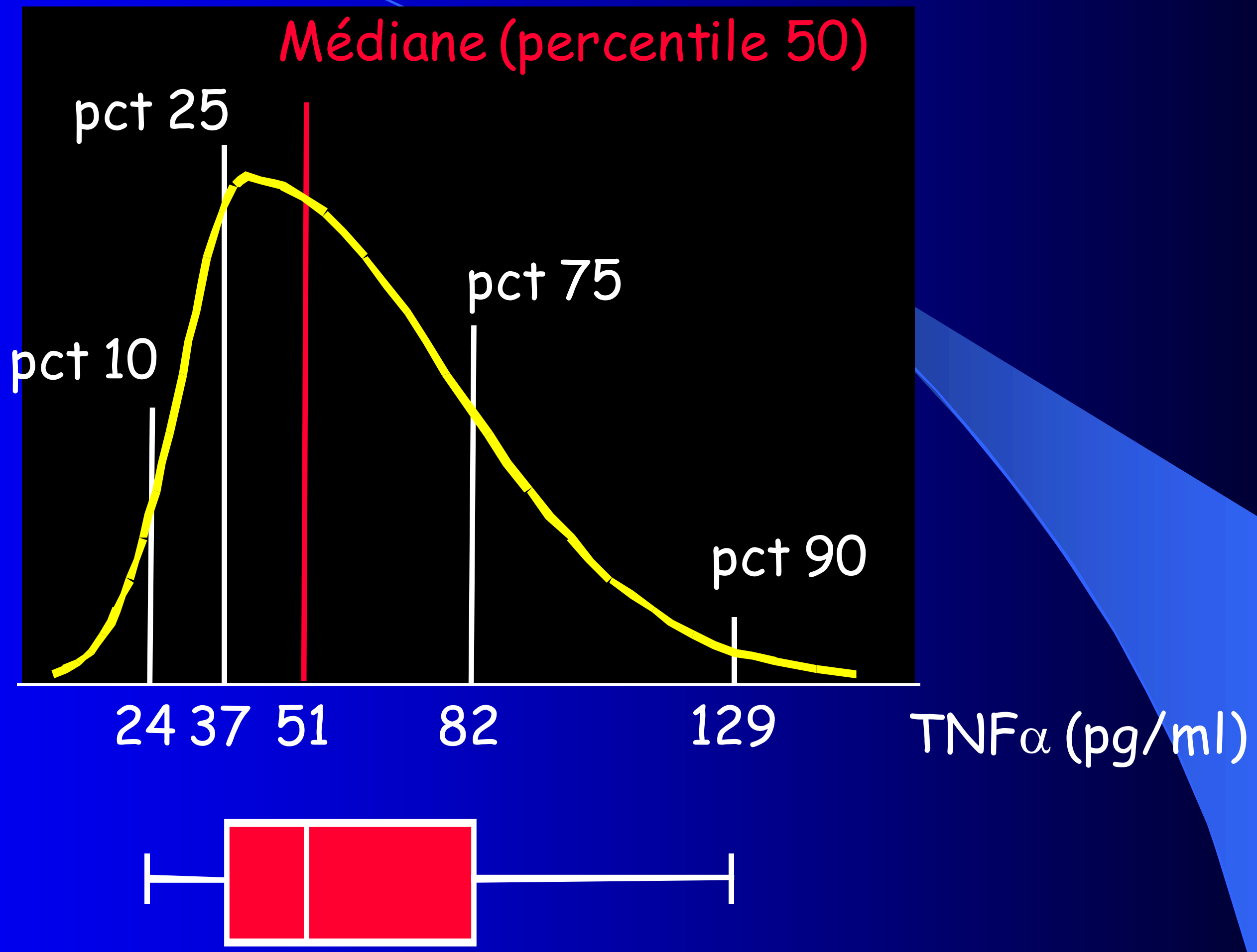


Coefficient d'asymétrie > 0.5
Skewness (3ème paramètre)



Coefficient d'asymétrie < - 0.5

Box plot (diagramme en boite)

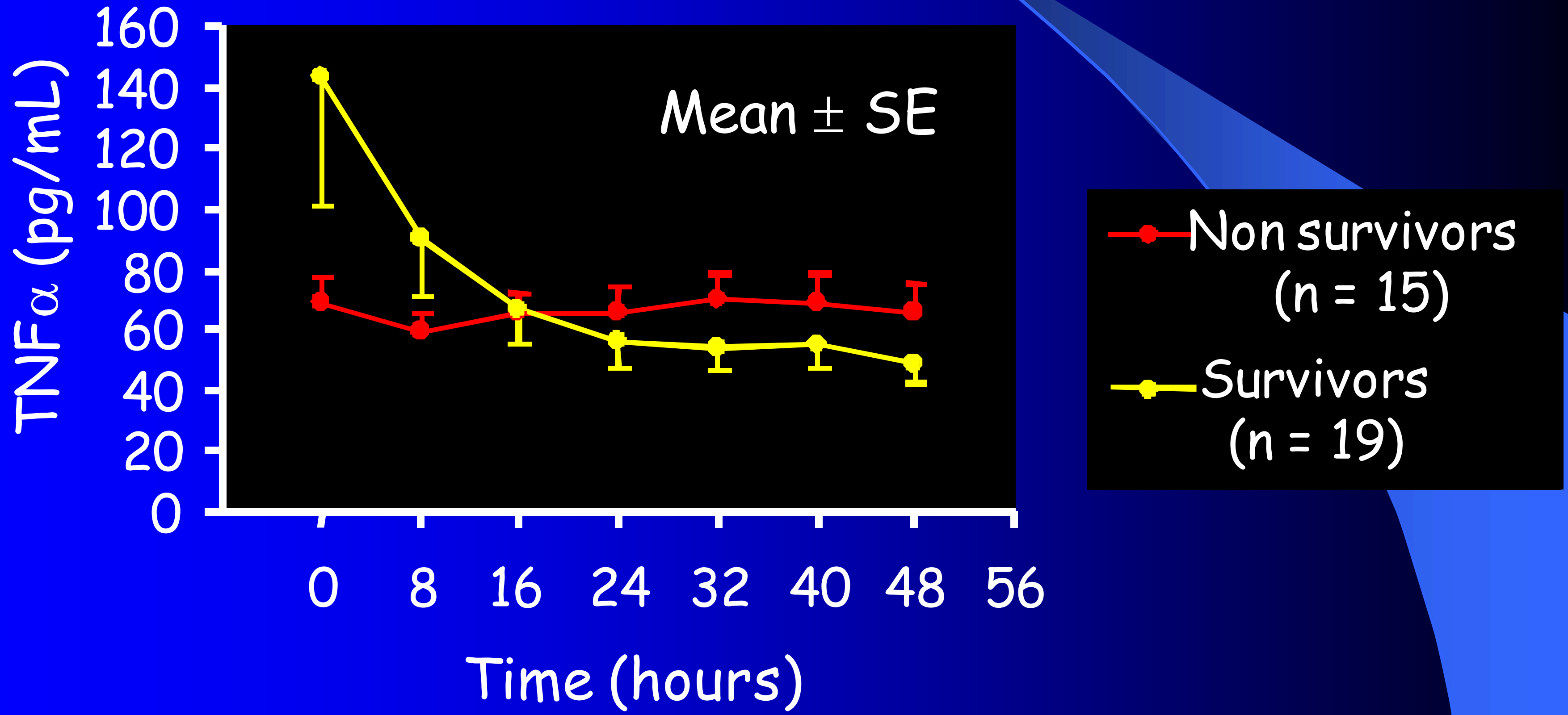


"box plot"

Tendance centrale: 51 pg/ml (médiane)

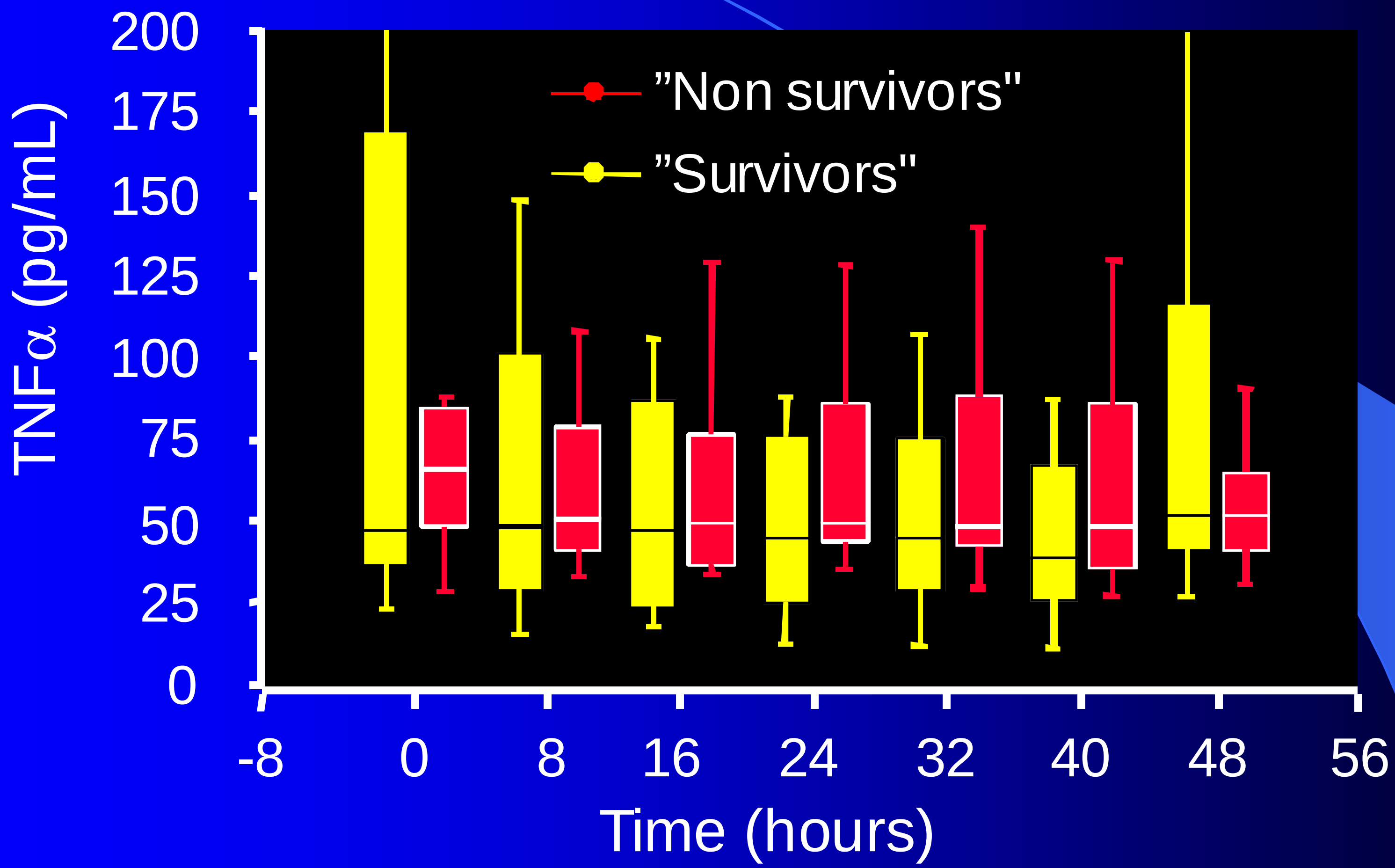
Dispersion: 37 à 82 (pct 25 à pct 75 = espace interquartile, IQR)

Représentation "paramétrique"



Conclusion: les non-survivants ont un taux initial de TNFα plus bas

Représentation «non paramétrique »



Conclusion: les survivants ont un taux initial de TNFα plus bas

Adapted from Appoloni O. et al Am J Med 2001;110:486-488

Conseil pour la lecture: les auteurs ont-ils analysé la structure des données?

Statistical Analysis

Descriptive continuous data were reported as mean±SD and compared with the Student *t* test. Categorical data were reported in percentage and tested by the Pearson χ^2 or Fisher exact test.

Stroke 1999;30:1402-1408

TABLE 1. Demographic and Clinical Characteristics at Admission

Characteristic	Treatment Surgical (n=99)	Treatment Endovascular (n=145)
Age, mean±SD, y	48.9±13.8	51.8±14.0
Time to admission,* mean±SD, h	19.5±29.1	25.8±36.3
Time to treatment,* mean±SD, h	43.1±37.3	48.4±40.8

- 38.7 à 77.7

Intervalle de prédiction à 95 % ?

Conseil pour la lecture: les auteurs ont-ils analysé la structure des données?

Statistical analysis

Descriptive statistics involved numbers and percentages of patients for categorical variables, and median with IQR for continuous variables. Univariate associations were tested by χ^2 or Kruskal-Wallis tests.

Lancet 2001;357:9-14

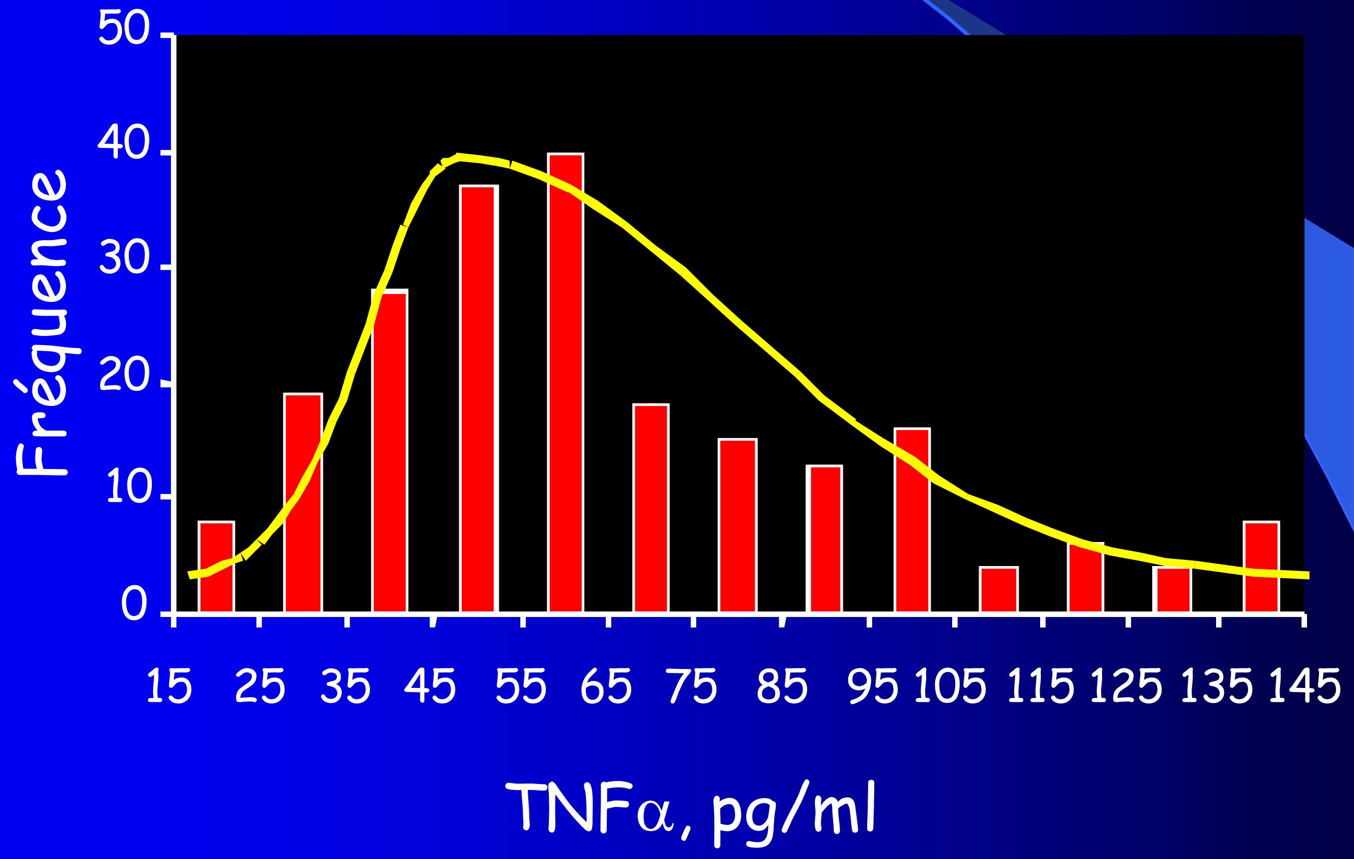
Statistical Analysis

Continuous variables, which were not normally distributed, are reported as medians and interquartile ranges. Categorical variables are reported as counts and percentages. Primary and secondary outcomes were binary, and the chi-square test or Fisher's exact test, as appropriate, was used to compare outcomes in the hypothermia and normothermia groups.

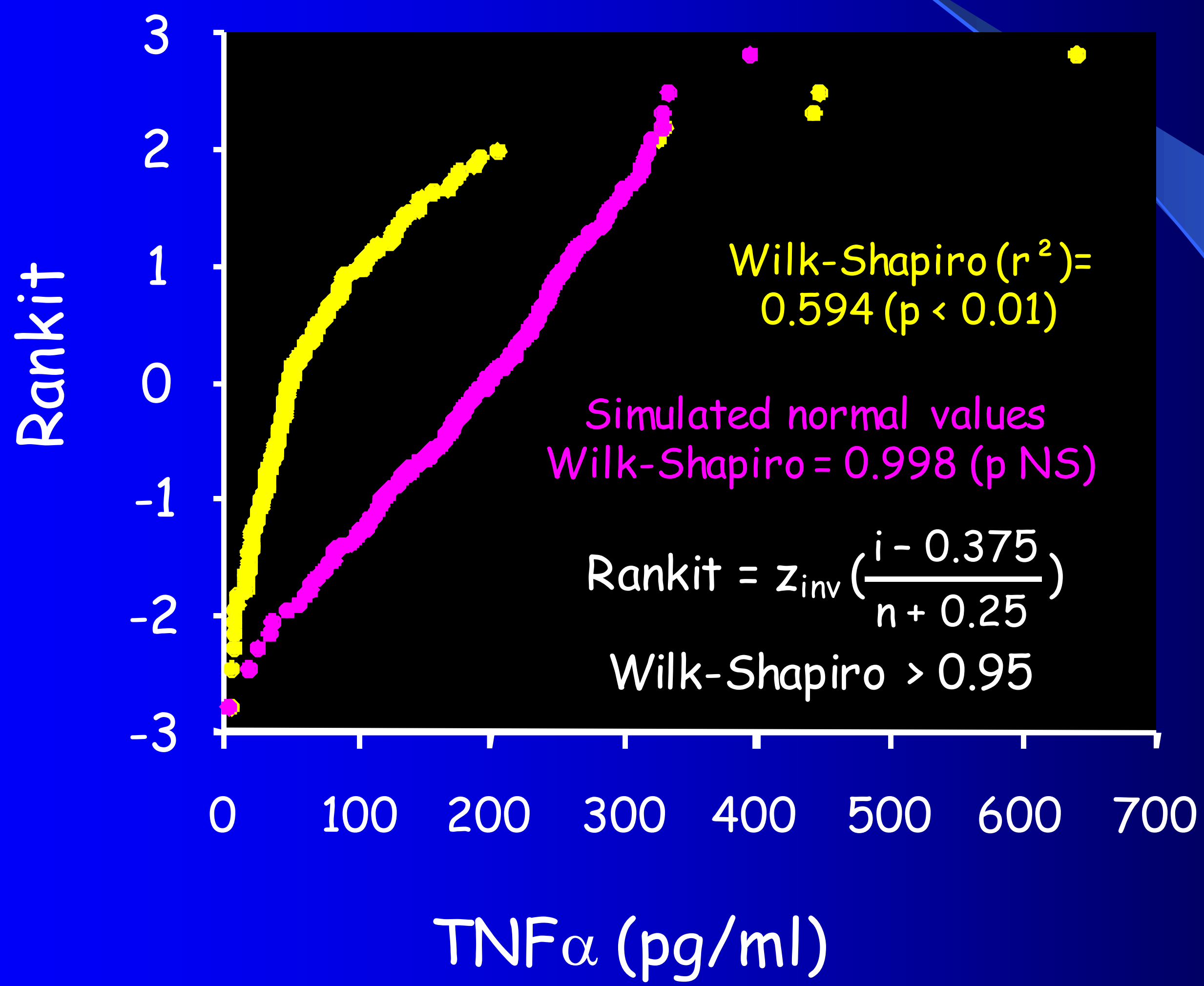
N Engl J Med 2002;549:556

- Comment évaluer la distribution des données?

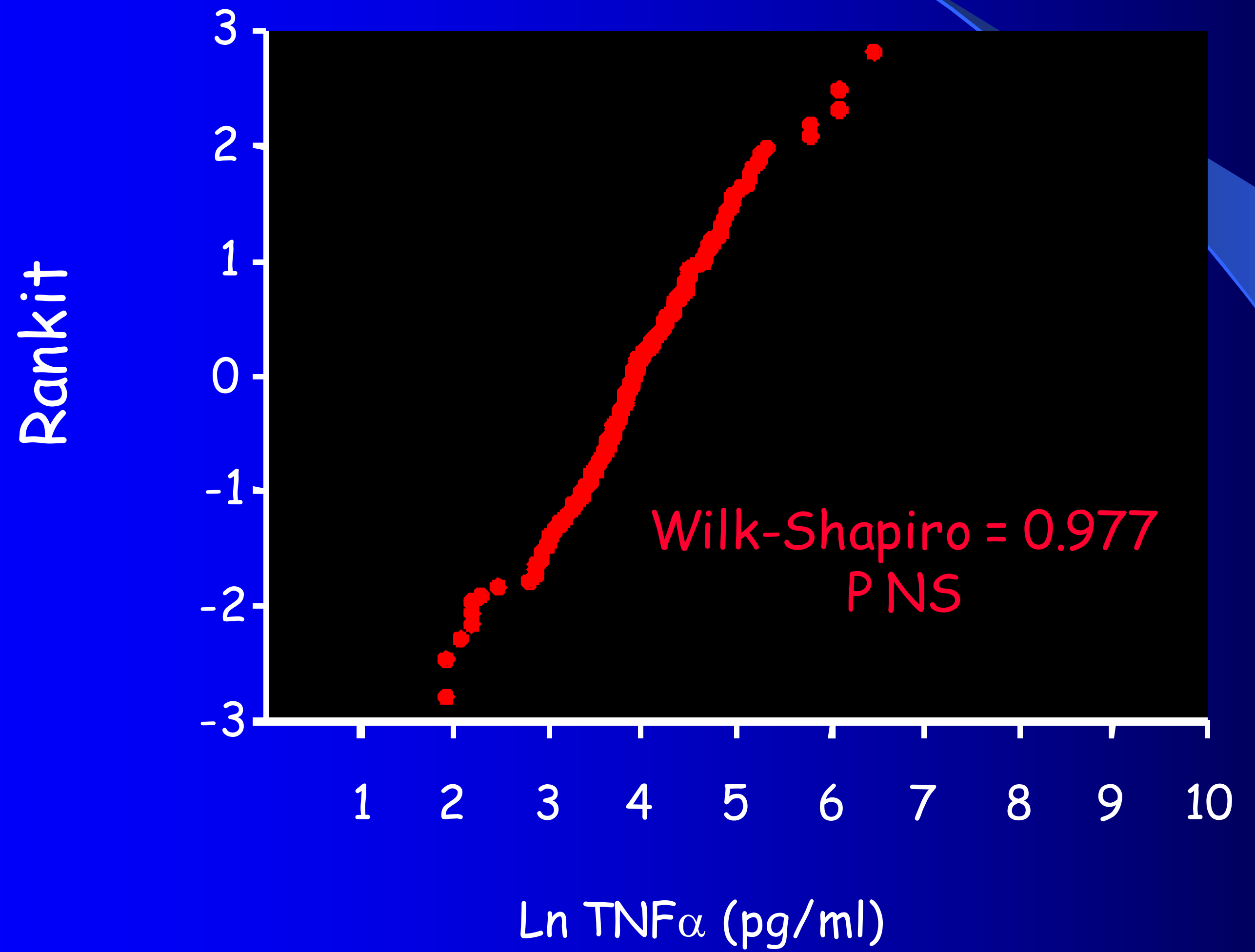
HISTOGRAMME



RANKIT PLOT et Test de Wilk - Shapiro



RANKIT PLOT: Tranformation logarithmique



Conseil pour la lecture: les auteurs ont-ils analysé la structure des données?

Statistical Analysis

Descriptive continuous data were reported as mean±SD and compared with the Student *t* test. Categorical data were reported in percentage and tested by the Pearson χ^2 or Fisher exact test.

Stroke 1999;30:1402-1408

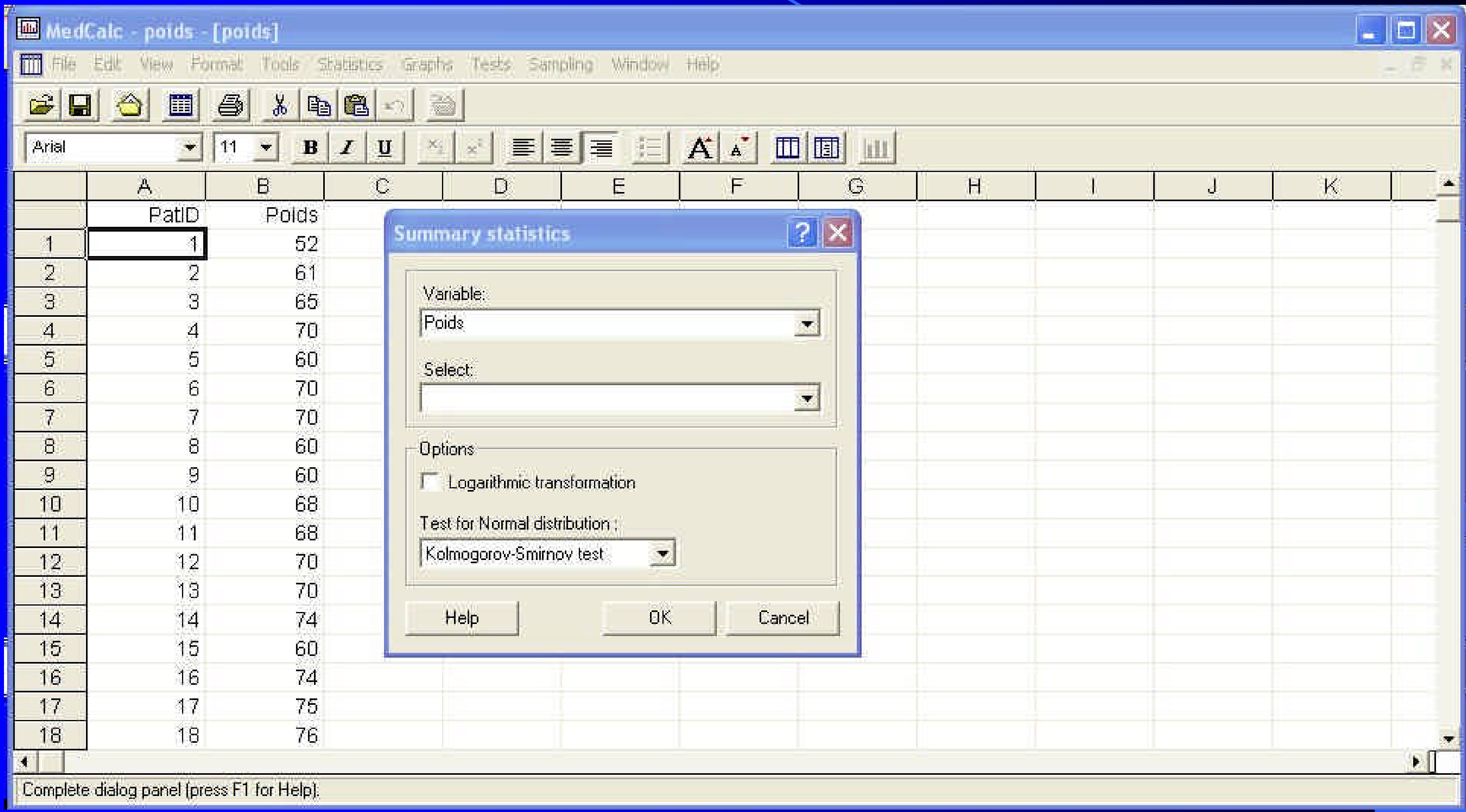
TABLE 1. Demographic and Clinical Characteristics at Admission

Characteristic	Treatment: Surgical (n=99)	Treatment: Endovascular (n=145)
Age, mean±SD, y	48.9±13.8	51.8±14.0
Time to admission,* mean±SD, h	19.5±29.1	25.8±36.3
Time to treatment,* mean±SD, h	43.1±37.3	48.4±40.8

- 38.7 à 77.7

Exemple

Patient	Poids	Patient	Poids	Patient	Poids	Patient	Poids
1	52	10	68	19	80	28	56
2	61	11	68	20	73	29	91
3	65	12	70	21	74	30	95
4	70	13	70	22	80	31	100
5	60	14	74	23	84	32	117
6	70	15	60	24	76	33	56
7	70	16	74	25	70	34	55
8	60	17	75	26	70	35	82
9	60	18	76	27	90		

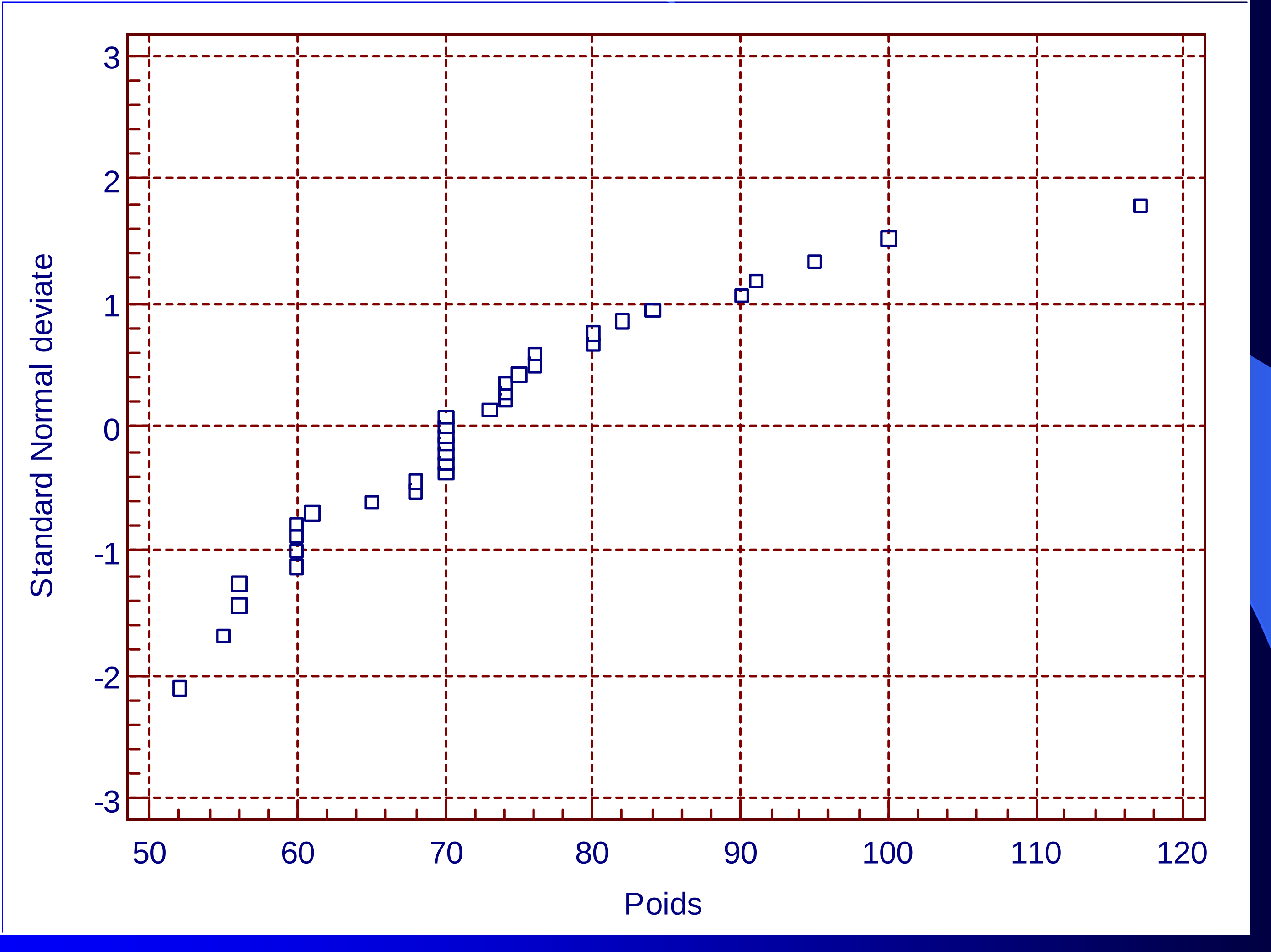

MedCalc - poids - [poids]
File Edit View Format Tools Statistics Graphs Tests Sampling Window Help
Arial
11
A B C D E F G H I J K
PatID Poids
1 1 52
2 2 61
3 3 65
4 4 70
5 5 60
6 6 70
7 7 70
8 8 60
9 9 60
10 10 68
11 11 68
12 12 70
13 13 70
14 14 74
15 15 60
16 16 74
17 17 75
18 18 76
Summary statistics
Variable:
Poids
Select:
Options
Logarithmic transformation
Test for Normal distribution :
Kolmogorov-Smirnov test
Help
OK
Cancel
Complete dialog panel (press F1 for Help).

Variable : Poids

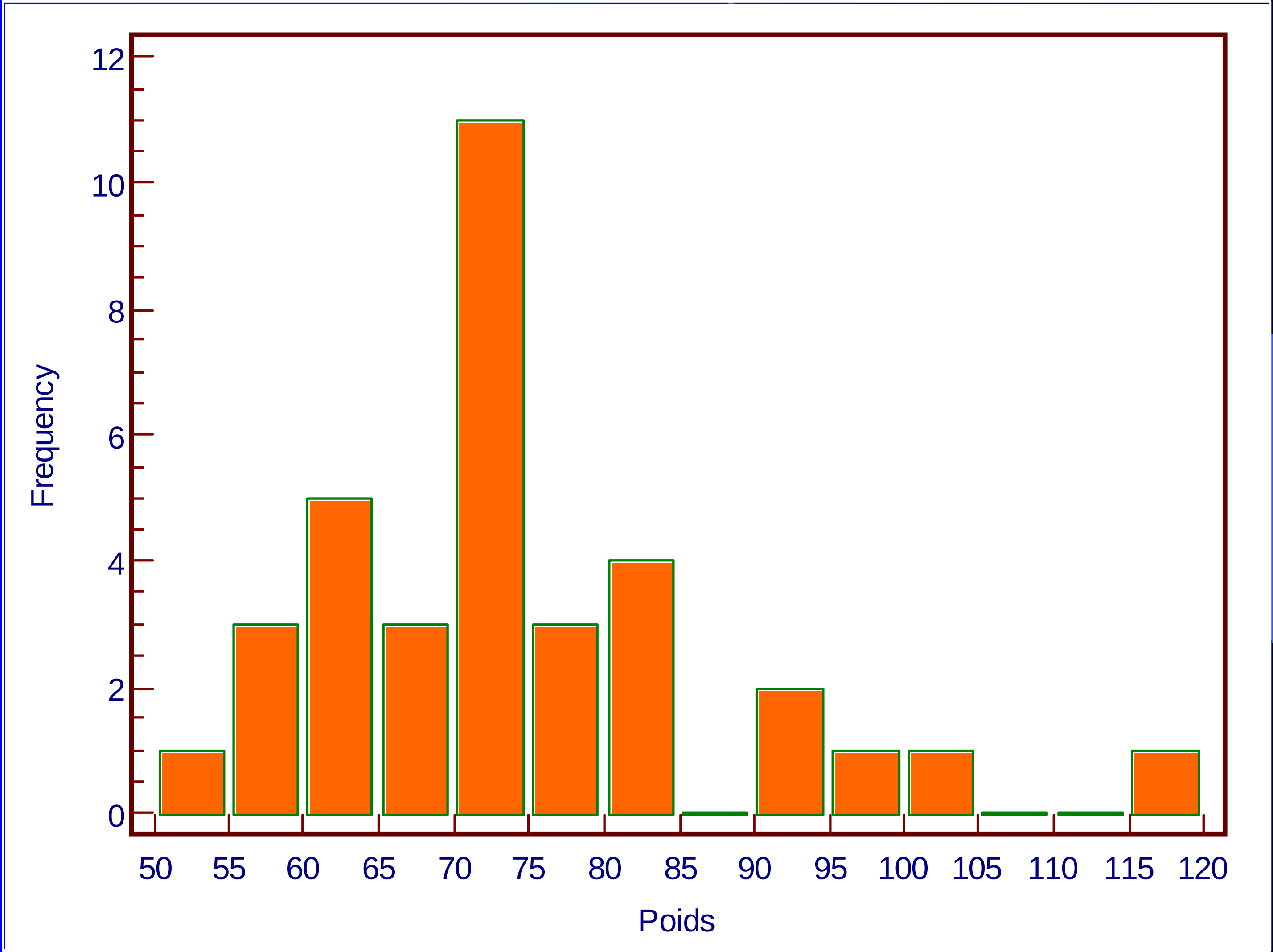
Sample size	=	35
Lowest value	=	52
Highest value	=	117
Arithmetic mean	=	72.9
95% CI for the mean	=	68.2 to 77.6
Median	=	70.0
95% CI for the median	=	68.0 to 75.6
Variance	=	189.0
Standard deviation	=	13.7
Standard error of the mean	=	2.3239

Kolmogorov-Smirnov test
for Normal distribution : accept Normality (p = 0.349)

Percentiles :			
2.5th =	53.1250	97.5th =	110.6250
5th =	55.2500	95th =	98.7500
10th =	56.0000	90th =	91.0000
25th =	62.0000	75th =	79.0000



Standard Normal deviate
3
2
1
0
-1
-2
-3
50
60
70
80
90
100
110
120
Poids

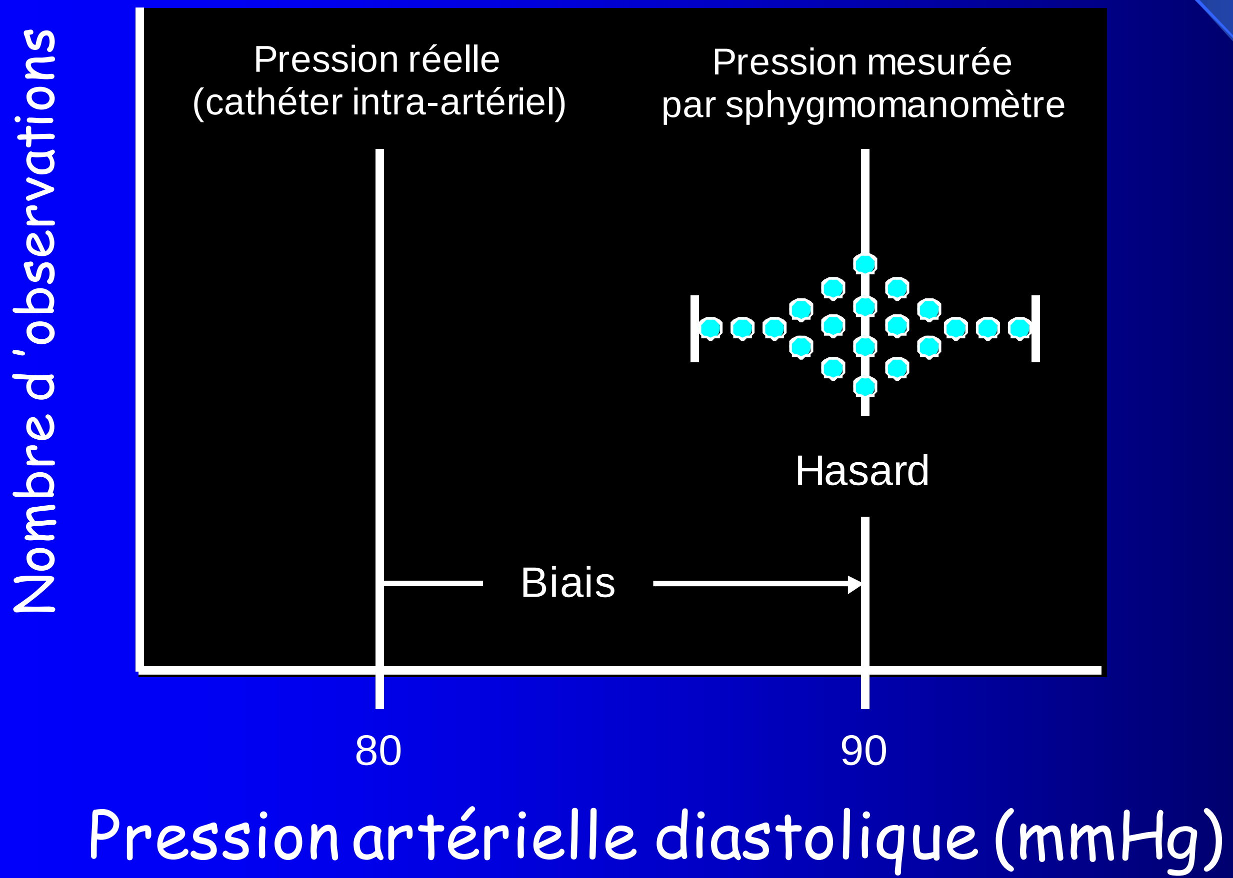

Frequency
12
10
8
6
4
2
0
50 55 60 65 70 75 80 85 90 95 100 105 110 115 120
Poids

- Comment mesurer la précision de l 'estimation d'une moyenne ou d'une proportion (degré d 'incertitude) ?

Biais et Hasard

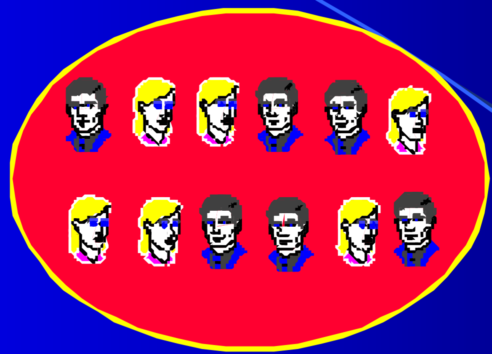
Effet observé
=
Effet réel
+
Erreur aléatoire ← Puissance
+
Erreur systématique ← Biais

Biais et Hasard



PRECISION DE L' ESTIMATION DE LA MOYENNE DE LA POPULATION: ERREUR STANDARD DE LA MOYENNE

population (μ, σ)



5 échantillons de taille n = 4

m_1 m_2 m_3 m_4 m_5

Les moyennes m_1, m_2, m_3, m_4 et m_5 sont distribuées normalement

avec une distribution dont les paramètres sont:

$$\begin{cases} \mu \\ SE = \dfrac{\sigma}{\sqrt{n}} \end{cases}$$

Estimés par:

$$\begin{cases} m \\ SE = \dfrac{s}{\sqrt{n}} \end{cases}$$

ERREUR STANDARD

$$SE(m) = \frac{SD}{\sqrt{n}} \qquad SE(p) = \frac{\sqrt{p\,(1-p)}}{\sqrt{n}}$$

Exemple: TAm = 110 mmHg
SD = 25 mmHg
n = 100
SE = 25 / 10 = 2.5 mmHg
n = 10000
SE = 25 / 100 = 0.3 mmHg

Exemple: p = 0.55 (55 %)
SD = $\sqrt{0.55 * 0.45}$ = 0.49
n = 100
SE = 0.49 / 10 = 0.049 (4.9 %)
n = 10000
SE = 0.49 / 100 = 0.0049 (0.5 %)

Si n augmente, SE diminue
et la puissance augmente

INTERVALLE DE CONFIANCE à 95 %

$IC_{95\,\%} = m \pm 2\ SE(m)$ $IC_{95\,\%} = p \pm 2\ SE(p)$

Exemple: TAm = 110 mmHg
n = 100
SE = 2.5 mmHg
IC 95 % = 105 à 115 mmHg
n = 10000
SE = 0.3 mmHg
IC 95 % = 109 à 111 mmHg

Exemple: p = 0.55 (55 %)
n = 100
SE = 0.049 (4.9 %)
IC 95 % = 0.45 à 0.65 (45 à 65 %)
n = 10000
SE = 0.0049 (0.5 %)
IC 95 % = 0.54 à 0.56
(54 à 56 %)

L'intervalle de confiance donne une estimation de la précision de la moyenne ou de la proportion calculée sur l 'échantillon de taille n

Odds ratio (Rapport des Cotes), Rapport de Risque et Différence de risque

$$\text{Odds ratio} = \frac{\text{Odds effet traitement A}}{\text{Odds effet traitement B}}$$

$$\text{Rapport de risque} = \frac{\text{Probabilité effet traitement A}}{\text{Probabilité effet traitement B}}$$

Différence de risque = Prob effet A – Prob effet B

Diagramme de Forest: interprétation du Odds Ratio et de son intervalle de confiance à 95%

$$IC\ 95\ \% = OR \pm 1.96\ SE$$

a	b
c	d

$$OR = \frac{a\,d}{b\,c}$$

$$SE(\ln(OR)) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

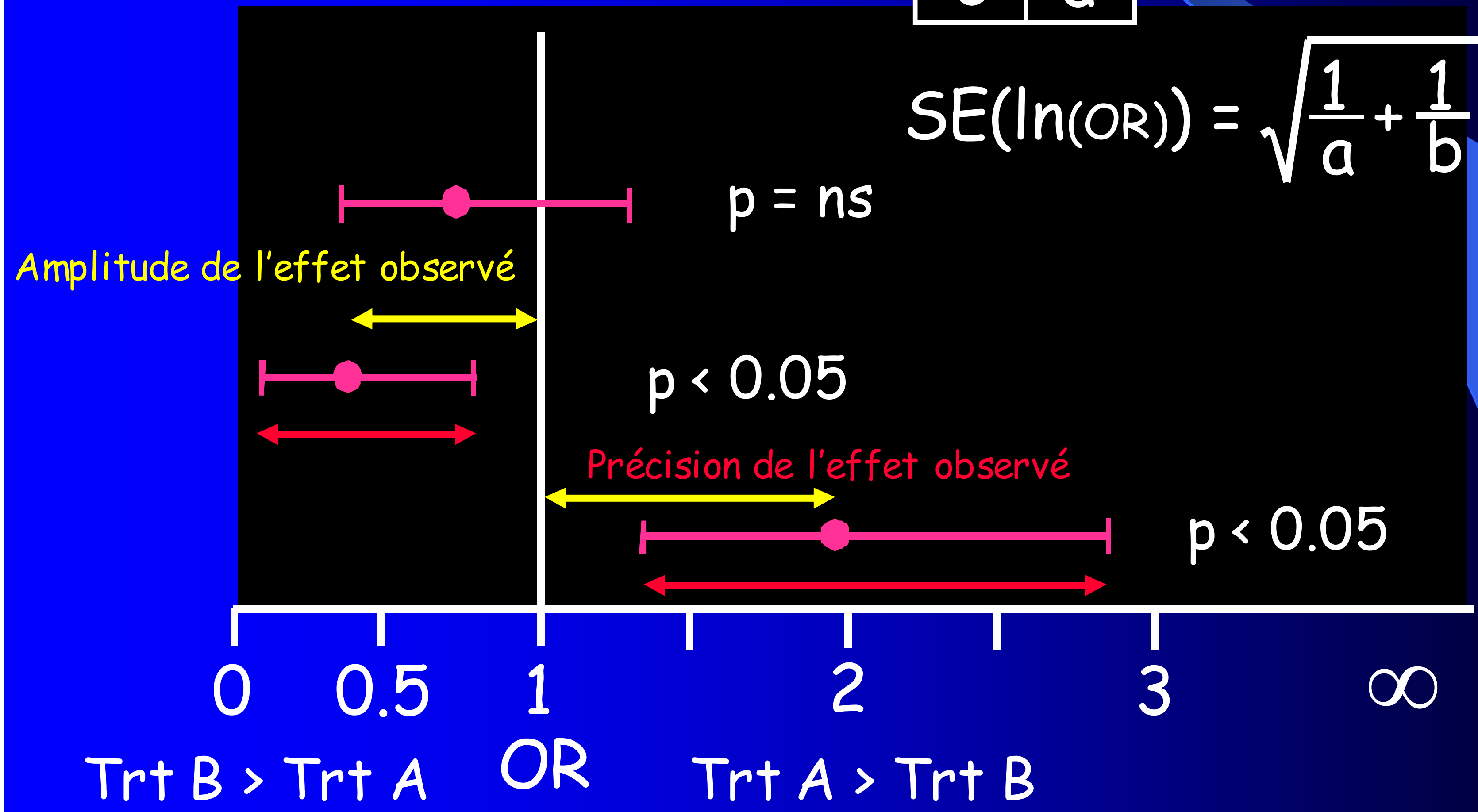


Diagramme de Forest: interprétation du Risque Relatif et de son intervalle de confiance à 95%

$$IC\ 95\ \% = RR \pm 1.96\ SE$$

a	b
c	d

$$RR = \frac{a/(a+b)}{c/(c+d)}$$

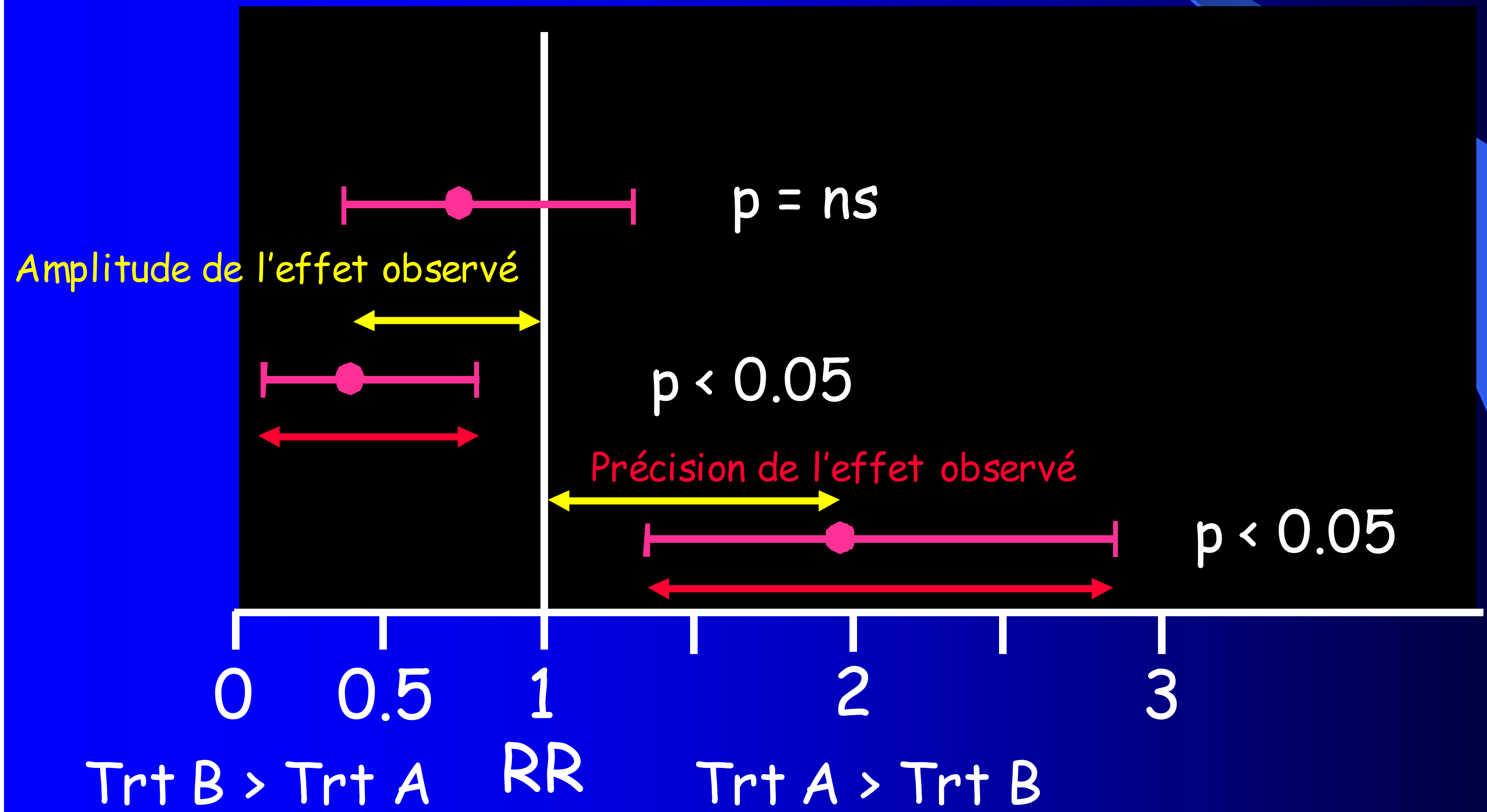
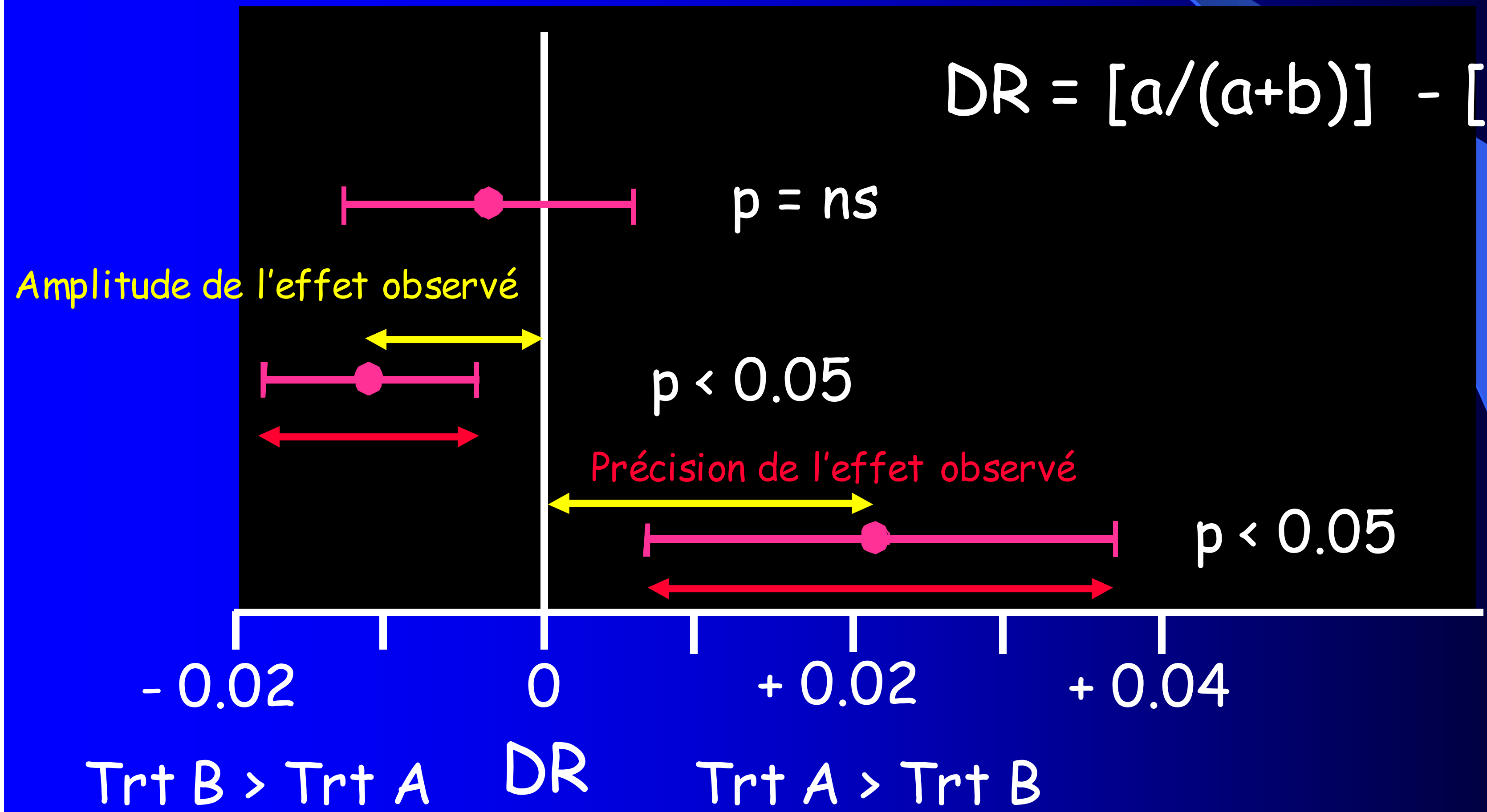


Diagramme de Forest: interprétation de la Différence de Risque et de son intervalle de confiance à 95%

IC 95 % = DR ± 1.96 SE

a	b
c	d



Exemple

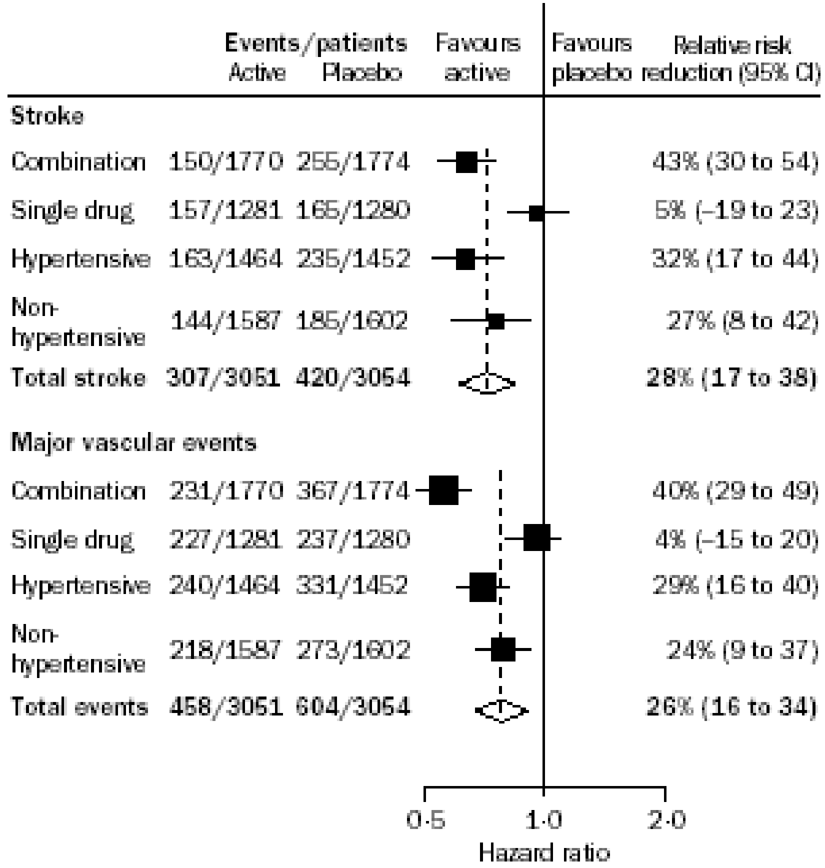


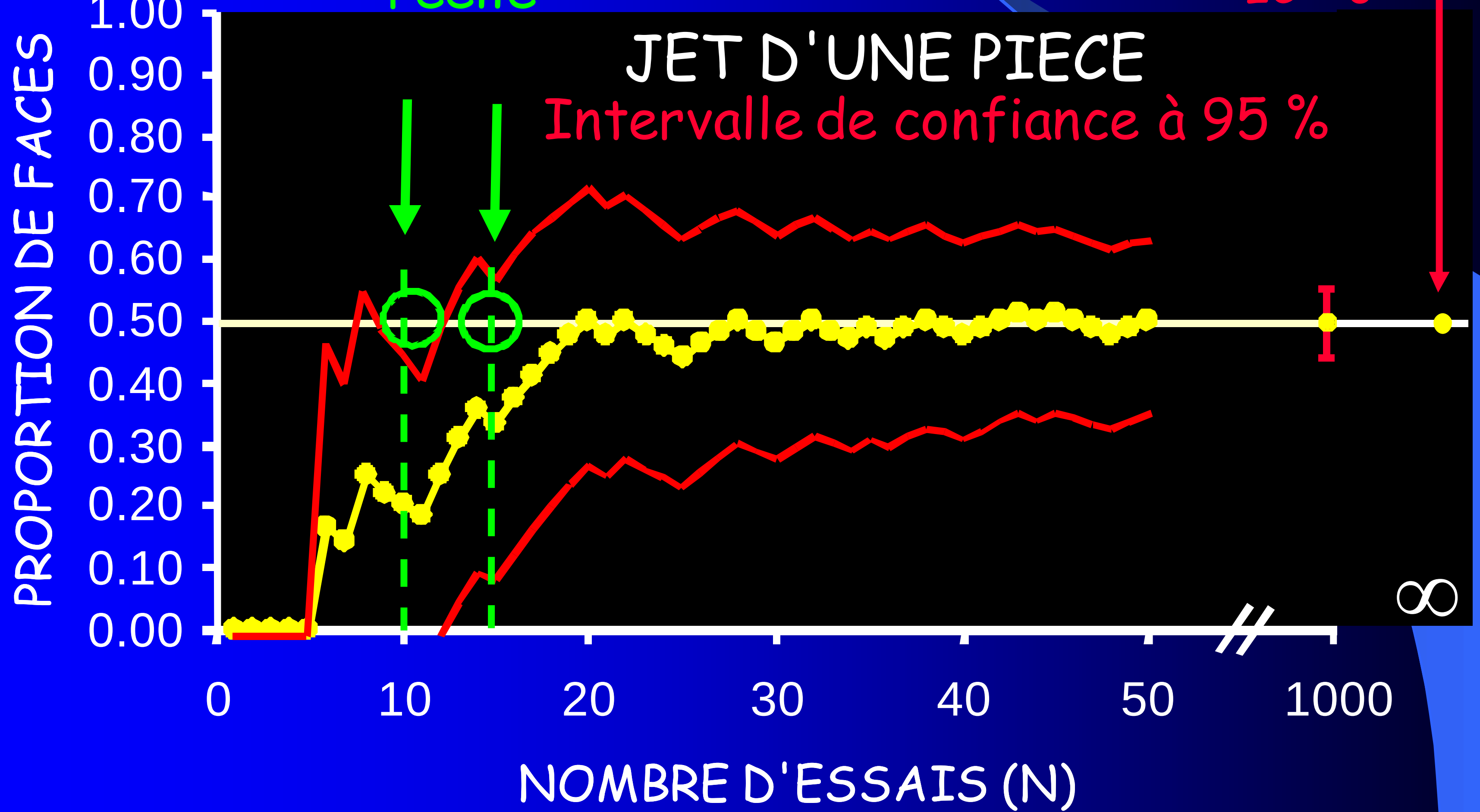
Figure 5: **Effects of study treatment on stroke and major vascular events in subgroups of patients**
Hazard ratios (and 95% CIs) for hypertensive and non-hypertensive subgroups standardised to study-wide proportions of patients for whom combination or single drug therapy was planned. p values for homogeneity (combination therapy vs single drug therapy) both <0·001. p values for homogeneity (hypertensive vs non-hypertensive) both >0·6. Conventions as in figure 4.

PROGRESS, Lancet 2001;358:1033-1041

Probabilité: intervalle de confiance

!! IC 95 %: contient (100 %) ou ne contient pas (0 %) la valeur réelle

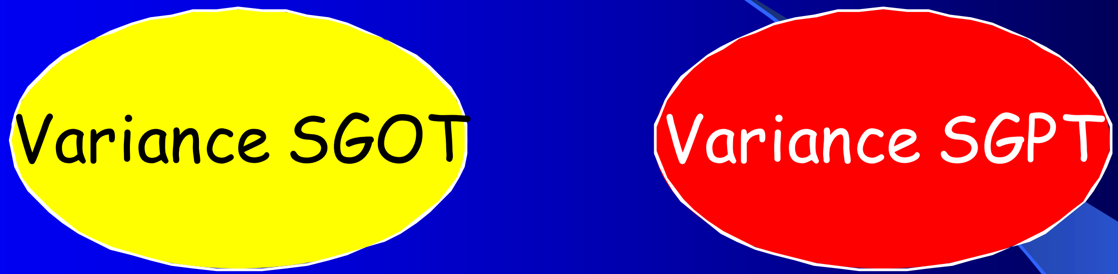
Il n'y a plus d'erreur
IC = 0



- Estimation du degré de liaison entre deux variables continues

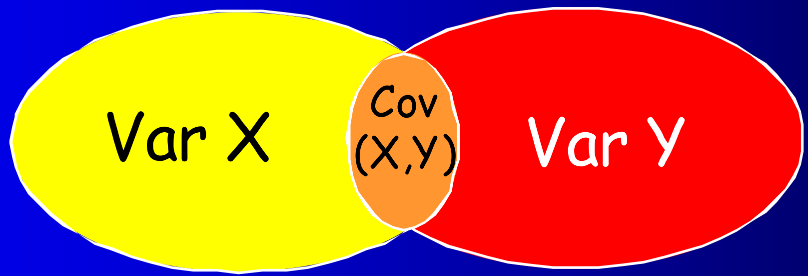
COVARIANCE

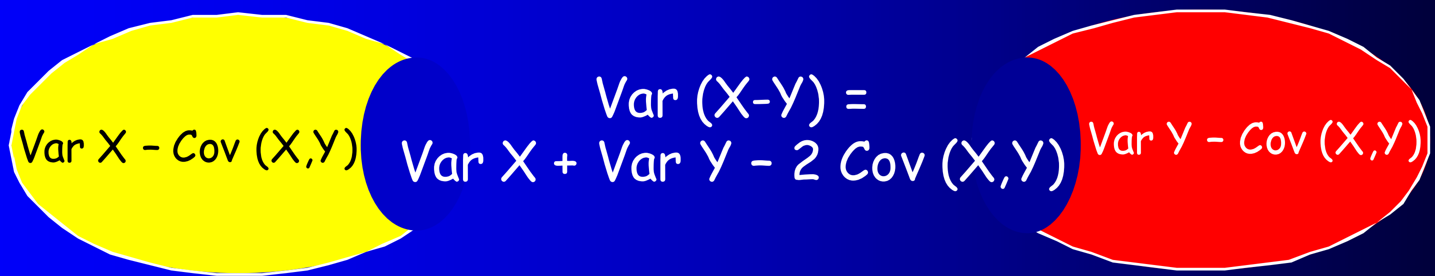
- Variables indépendantes (Cov(X,Y) = 0):



Var (X-Y) = Var X + Var Y

- Variables non indépendantes (Cov(X, Y) ≠ 0:





COVARIANCE

Covariance: lien entre deux variables (Cov(X,Y)):

	X = SGOT		Y = SGPT	(X i - moy.)	(Yi - moy.)	(X i - moy.) (Yi - moy.)	
	12		52	-5.8	11.4	-66.1	
	15		45	-2.8	4.4	-12.3	
	17		42	-0.8	1.4	-1.1	
	22		39	4.2	-1.6	-6.7	
	23		25	5.2	-15.6	-81.1	
Somme	89		203	0	0	-167.4	Produit croisé
n	5		5			4	ddl
Moyenne	17.8		40.6			-41.9	Cov (X,Y)

Si les variables sont indépendantes, Cov = 0

STRUCTURE DE LA COVARIANCE

Somme des produits croisés

$$\mathrm{Cov}\,(x, y) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})\,(Yi - \bar{Y})}{n - 1}$$

Degré de liberté
(n = nombre de paires)

COMPARISON DE DEUX MOYENNES

(échantillons indépendants)

Test t non pairé de Student

$$t = \frac{|m_x - m_y|}{\sqrt{\frac{1}{n}(SD_x^2 + SD_y^2)}} = \frac{22.8}{\sqrt{\frac{1}{5}(21.7 + 99.3)}} = 4.63$$

Avec 4 + 4 = 8 ddl

Valeur de p = 0.0017

L'expérimentateur décide que la covariance = 0

COMPARISON DE DEUX MOYENNES
(échantillons non indépendants)

Test t pairé de Student

$$t = \frac{|\ m_x - m_y\ |}{\sqrt{\frac{1}{n}\left(SD_x^2 + SD_y^2 - 2\ \mathrm{Cov}\ (x, y)\right)}}$$

$$t = \frac{22.8}{\sqrt{\frac{1}{5}\left(21.7 + 99.3 - 2\ (-41.85)\right)}} = 3.56$$

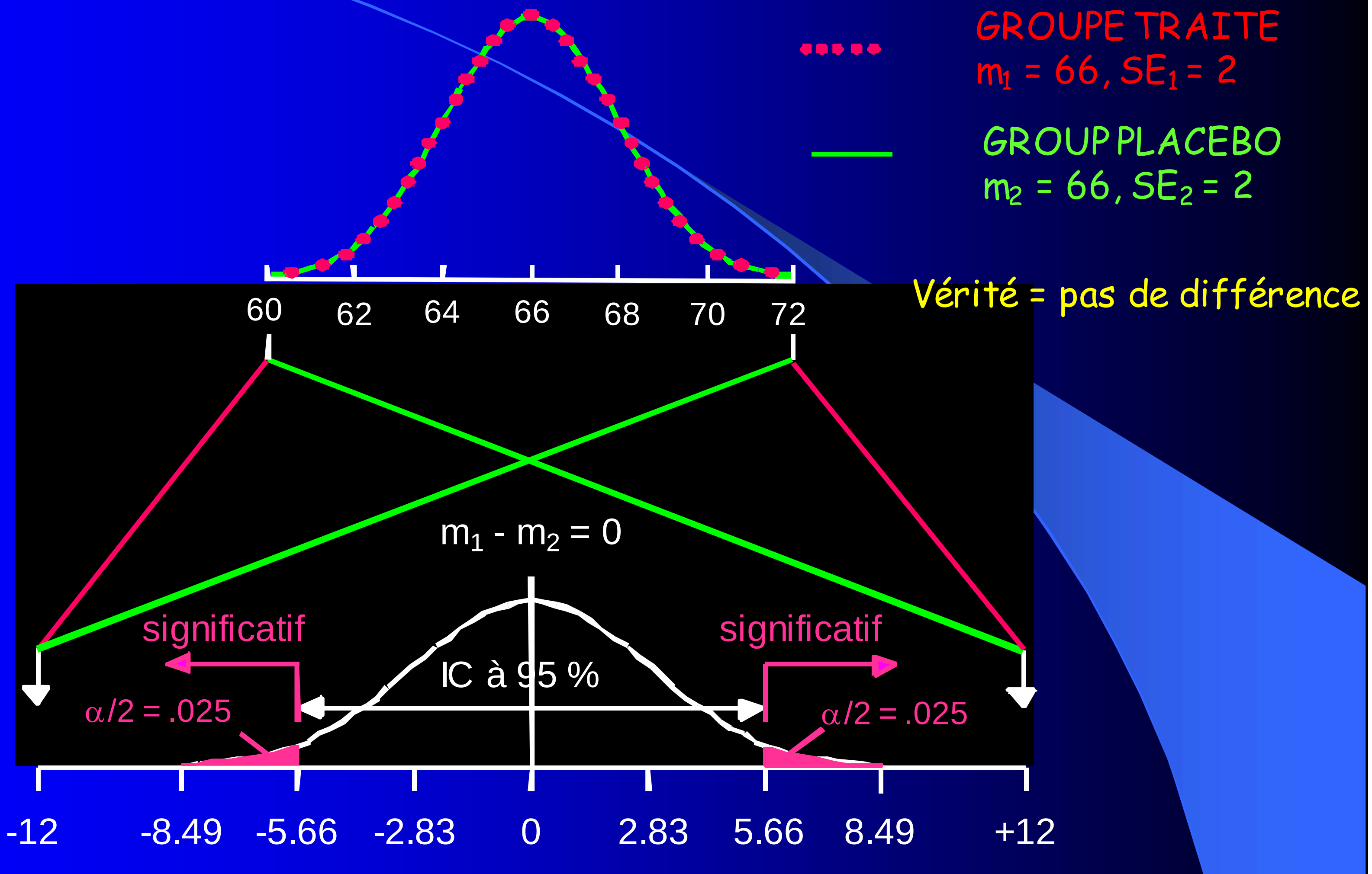
avec 4 + 4 – 4 = 4 ddl

Valeur de p = 0.0235

L'expérimentateur décide de tenir compte de la covariance

- Le hasard peut-il créer une différence entre deux groupes?

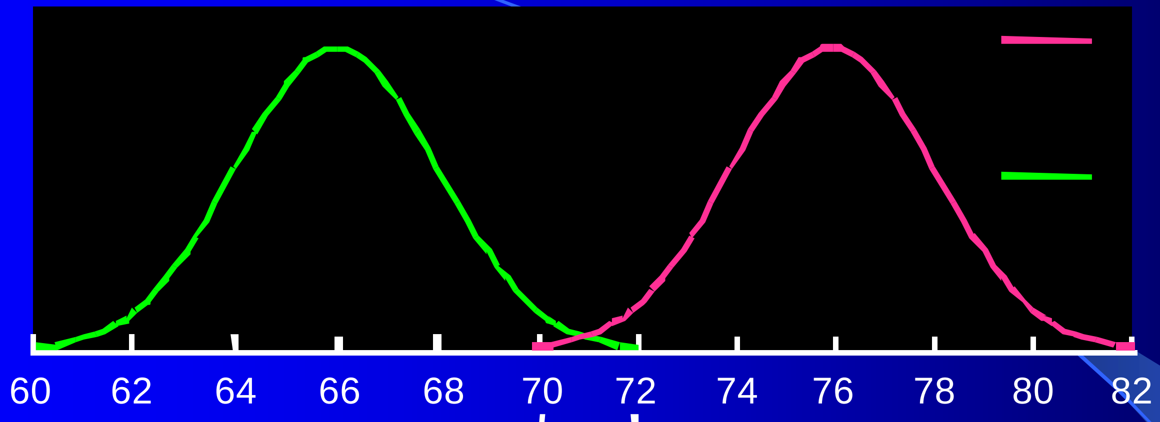
HYPOTHESE NULLE



$$SE_{m1-m2} = \sqrt{SE^2_1 + SE^2_2} = 2.83$$

En l'absence de différence, les fluctuations autour de zéro due au hasard (fluctuations d'échantillonnage) atteignent le niveau significativité dans 5 % des cas (erreur de type I ou erreur α)

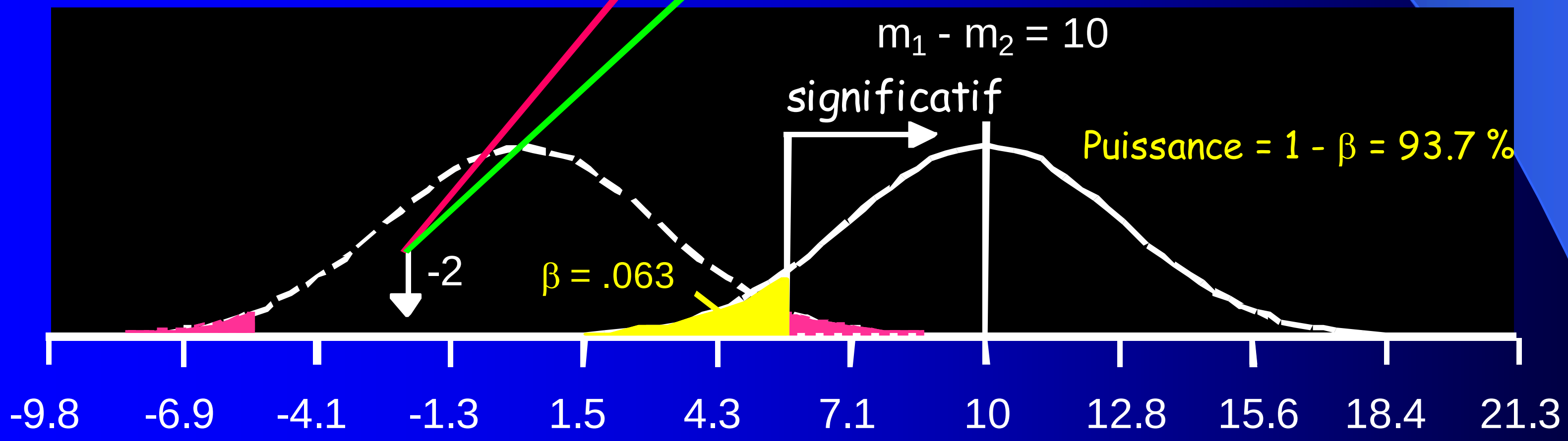
HYPOTHESE ALTERNATIVE et ERREUR de TYPE II (β)



GROUPE TRAITE
$m_1 = 76$, $SE_1 = 2$

GROUPE PLACEBO
$m_2 = 66$, $SE_2 = 2$

Ha est vraie: différence = 10



$$SE_{m1\text{-}m2} = \sqrt{SE^2_1 + SE^2_2} = 2.83$$

En présence d'une différence, les fluctuations autour de la différence moyenne (= 10) liées au hasard seul (fluctuations d'échantillonnage) tombent en dessous du seuil de significativité dans 6.3 % des cas (erreur de type II ou erreur or β)

Types d'erreurs

La réalité: placebo vs traitement

	$T \neq P$ Ha	$P = T$ Ho
$p < 0.05$	correct	Erreur α
p NS	Erreur β	correct

Test statistique sur 2 échantillons

Principe des tests statistiques: l'hypothèse nulle

- Le principe général des tests statistiques dits tests d'hypothèse repose sur la formulation d'une hypothèse nulle (appelée hypothèse privilégiée) que l'on cherche à rejeter au profit de l'hypothèse alternative.

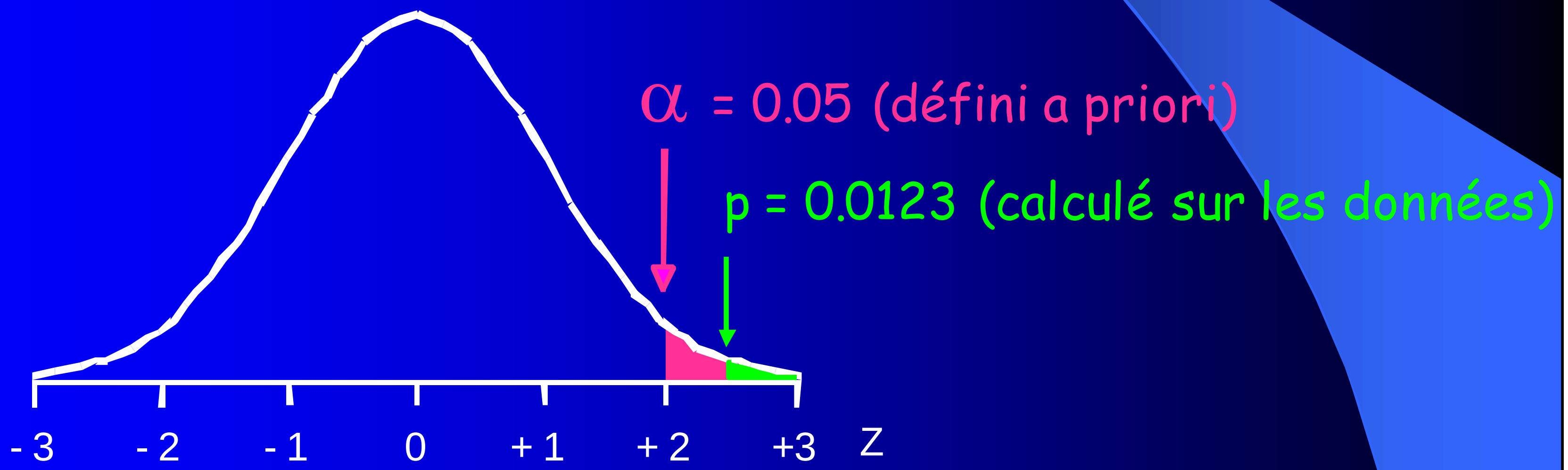
- Rejeter l'hypothèse nulle, c'est accepter qu'il existe une différence significative (non produite par le hasard seul) entre les deux éléments comparés (par ex.: deux traitements).

Signification statistique (p)

- La valeur de p mesure la probabilité que le hasard explique la différence observée ou une différence plus grande, s'il n'y avait pas de différence (sous l'hypothèse nulle)

- Si p est inférieur au taux α (erreur de type I, 5 %), la différence est statistiquement significative et nous rejetons l'hypothèse nulle pour accepter l'hypothèse alternative (il y a une différence)

Valeur de p (p-value) et taux d'erreur de type I (α).



- Tests statistiques élémentaires

Tests statistiques

- Comparaisons des variables continues:
  - Les tests paramétriques (moyennes, variances):
    - Différences entre deux moyennes: test t de Student
    - Le rapport entre deux variances: le test de F
  - Les tests non paramétriques (rangs)
- Comparaisons des variables discrètes
    - Différence entre deux pourcentages: test de z
    - Différences entre deux répartitions: test du Chi ²

TEST t de STUDENT

(Comparaison de deux échantillons indépendants)

EXEMPLE:

But de l'étude:

étudier l'effet d'un β-bloquant sur la pression artérielle systolique.

Design:

10 patients hypertendus randomisés en deux groupes de 5 patients: un groupe recevant le placebo et un groupe recevant le β-bloquant.

TEST t de STUDENT

(Comparaison de deux échantillons indépendants)

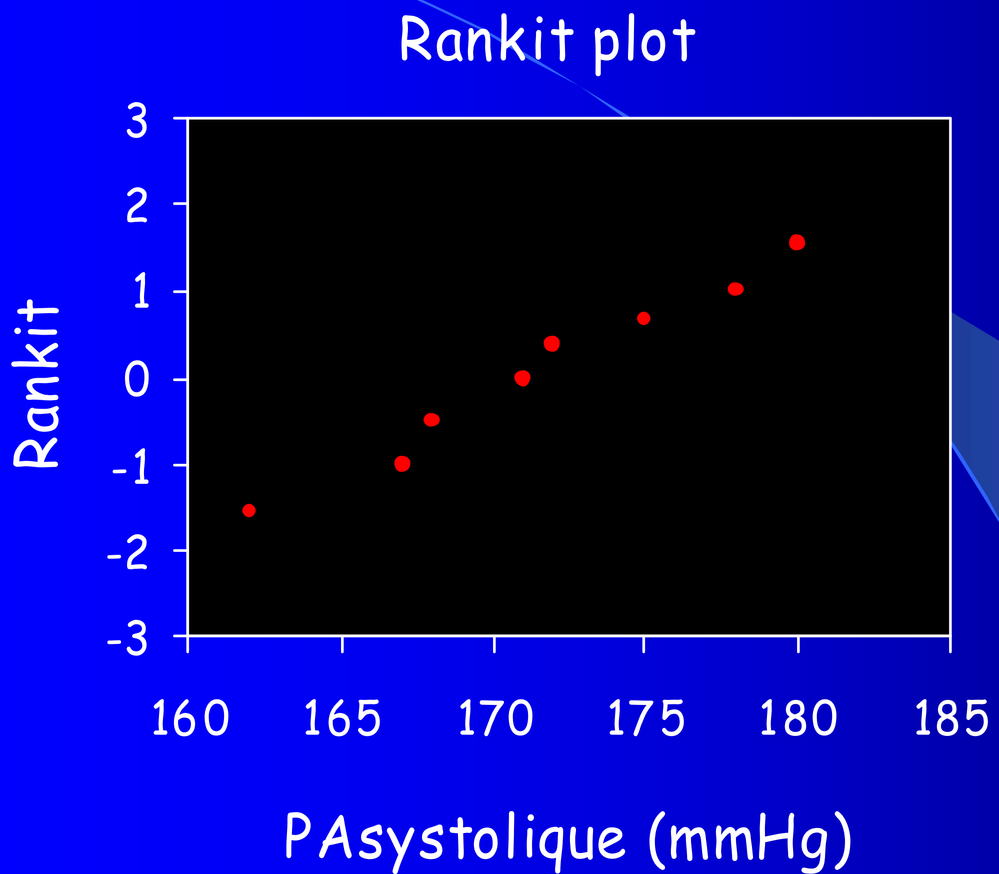
Resultats:

		Pression artérielle systolique (mmHg)	
		Groupe 1 Placebo	Groupe 2 β bloquant
Patients	1	180	167
	2	175	171
	3	168	162
	4	171	168
	5	178	172
Moyenne		174.4	168.0
SD		4.9	3.9

TEST t de STUDENT

(Comparaison de deux échantillons indépendants)

1) Vérifier la normalité des données:



Wilk-Shapiro = 0.9698

$$\text{Rankit} = z_{inv}\left(\frac{i - 0.375}{n + 0.25}\right)$$

TEST t de STUDENT

(Comparaison de deux échantillons indépendants)

2) Vérifier l'égalité des variances:

$$F = \frac{24.30}{15.52} = 1.57 \quad \text{avec} \quad \begin{matrix} ddl = 4 \\ ddl = 4 \end{matrix} \qquad p = 0.337$$

Si variances inégales → approximation de Satterthwaite

$$t = \frac{|m_1 - m_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad \text{avec d.d.l.} = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{\left[\frac{s_1^2}{n_1}\right]^2}{(n_1 - 1)} + \frac{\left[\frac{s_2^2}{n_2}\right]^2}{(n_2 - 1)}}$$

TEST t de STUDENT

(Comparaison de deux échantillons indépendants)

3) Calcul de la variance poolée:

$$S_p = \sqrt{\frac{(n_1 - 1)\,S_1^2 + (n_2 - 1)\,S_2^2}{n_1 + n_2 - 2}}$$

$$s = \sqrt{\frac{4 \times 4.93^2 + 4 \times 3.94^2}{8}} = 4.46$$

TEST t de STUDENT

(Comparaison de deux échantillons indépendants)

3) Calcul du test de t:

$$t = \frac{|m_1 - m_2|}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{6.4}{4.46\sqrt{\frac{1}{5} + \frac{1}{5}}} = 2.27$$

ddl = 5 + 5 - 2 = 8

p = 0.053 (NS)

TEST non paramétrique de Mann-Whitney

(Comparaison de deux échantillons indépendants)

Résultats:

		Pression artérielle systolique (mmHg)	
		Groupe 1 Placebo	Groupe 2 β bloquant
Patients	1	180	167
	2	175	171
	3	168	162
	4	171	168
	5	178	172
Moyenne SD		174.4 4.9	168.0 3.9

TEST non paramétrique de Mann-Whitney

(Comparaison de deux échantillons indépendants)

1) Calcul des rangs:

Groupe	Valeurs ordonnées	Rang	Valeurs identiques (« ties »)
2	162	1	
2	167	2	
1	168	3	3.5
2	168	4	3.5
1	171	5	5.5
2	171	6	5.5
2	172	7	
1	175	8	
1	178	9	
1	180	10	

TEST non paramétrique de Mann-Whitney

(Comparaison de deux échantillons indépendants)

2) Remplacer les données par leurs rangs:

		Pression artérielle systolique (mmHg)	
		Groupe 1 Placebo	Groupe 2 β bloquant
Patients	1	10	2
	2	8	5.5
	3	3.5	1
	4	5.5	3.5
	5	9	7
Somme des rangs		36	19

Wx Wy

TEST non paramétrique de Mann-Whitney

(Comparaison de deux échantillons indépendants)

3) Calcul du test de Mann-Whitney:

Comparisons	W stat	p (1 tail)	p (2 tails)
Group 1 & Group 2	36 vs 19	0.0476	0.0952

W = rank sum test

TEST non paramétrique: test U

(Comparaison de deux échantillons indépendants)

1) Calcul des rangs:

Groupe	Valeurs ordonnées	U(Y,Xi)	U(X,Yi)
2	162		0
2	167		0
1	168	2,5	
2	168		0,5
1	171	3,5	
2	171		1,5
2	172		2
1	175	5	
1	178	5	
1	180	5	

Ux = 21 Uy = 4

Robust rank-order

TEST non paramétrique: test U

(Comparaison de deux échantillons indépendants)

3) Calcul du test U:

Comparisons	U stat	p (1 tail)	p (2 tails)
Group 1 & Group 2	21 vs 4	0.0476	0.0947

U = robust rank-order test

TEST t de STUDENT

(Comparaison de deux échantillons pairés)

EXEMPLE:

But de l'étude:

étudier l'effet d'un β-bloquant sur la pression artérielle systolique.

Design:

5 patients hypertendus

mesure de la pression artérielle avant et après la prise de β-bloquant.

TEST t de STUDENT

(Comparaison de deux échantillons pairés)

Resultats:

		Pression artérielle systolique (mmHg)		
		Base	β bloquant	Différence
Patients	1	180	167	13
	2	175	171	4
	3	168	162	6
	4	171	168	3
	5	178	172	6
Moyenne		174.4	168.0	6.4
SD		4.9	3.9	3.9

TEST t de STUDENT

(Comparaison de deux échantillons pairés)

1) Calcul du test de t pairé:

$$t = \frac{|m_{dif}|}{\sqrt{\frac{s^2_{dif}}{n}}} = \frac{6.4}{\sqrt{\frac{3.9^2}{5}}} = 3.66 \quad \text{avec d.d.l.} = n - 1 = 4$$

p = 0.0216

TEST non paramétrique de Wilcoxon des rangs signés

(Comparaison de deux échantillons pairés)

1) Calculer les différences de rangs

		Pression artérielle systolique (mmHg)		
		Base	β bloquant	Différence
Patients	1	10	2	+8
	2	8	5.5	+2.5
	3	3.5	1	+2.5
	4	5.5	3.5	+2
	5	9	7	+2
Somme des rangs		36	19	+17

TEST non paramétrique de Wilcoxon des rangs signés

(Comparaison de deux échantillons pairés)

2) Calculer le test de Wilcoxon:

Comparisons	Sum negative & positive ranks	p (1 tail)	p (2 tail)
Mean 1 & Mean 2	0 vs 17	0.0312	0.0591

2) Calculer le test du signe:

Comparisons	Number of negative & positive differences	p (1 tail)	p (2 tail)
Mean 1 & Mean 2	0 vs 5	0.031	0.062

VARIABLES CONTINUES	Distribution gaussienne Tests paramétriques	Distribution non-gaussienne Tests non paramétriques
2 moyennes (groupes indépendants)		
- variances égales	Test t non pairé	Mann-Withney U test Median test
- variances inégales	Test t non pairé (modification de Satterthwaite)	
2 moyennes (groupes pairés)		
	Test t pairé	Sign test Wilcoxon signed rank test

- Comparaisons des variables discrètes

Y a-t-il une relation entre un facteur de risque et la maladie? (non indépendance)

		Maladie		
		Présente (cancer bronchique)	Absente (pas de cancer)	
Facteur de Risque	Présent (fumeur)	15	85	100
	Absent (non fumeur)	3	197	200
		18	282	300

Ce que le hasard aurait mis en l'absence de relation entre le facteur de risque et la maladie (indépendance)

		Maladie		
		Présente (cancer bronchique)	Absente (pas de cancer)	
Facteur de Risque	Présent (fumeur)	$18 \times \frac{100}{300} = 6$	$282 \times \frac{100}{300} = 94$	100
	Absent (non fumeur)	$18 \times \frac{200}{300} = 12$	$282 \times \frac{200}{300} = 188$	200
Les totaux marginaux sont fixés		18	282	300

Le test du χ^2 (Chi²) de Pearson

		Maladie		
		Présente (cancer bronchique)	Absente (pas de cancer)	
Facteur de Risque	Présent (fumeur)	Obs: 15 Exp: 6	Obs: 85 Exp: 94	100
	Absent (non fumeur)	Obs: 3 Exp: 12	Obs: 197 Exp: 188	200
		18	282	300

$\chi^2 = \sum (Obs - Exp)^2 / Exp$

$= (15-6)^2/6 + (85-94)^2/94 + (3-12)^2/12 + (197-188)^2/188$

$= 13.5 + 0.86 + 6.75 + 0.43$

$= 21.54$

avec 1 ddl $(r-1) \times (c-1) \rightarrow$ $p = 0.00001$

Test du χ^2 (Chi²)

- Si une valeur attendue est inférieure à 5:
  - Correction de Yates
- Si une valeur attendue est inférieure à 3:
  - Test exact de Fisher

Test du χ^2 pour mesures répétées (Test de Mac Nemar)

- Exemple: mesure des difficultés à l'endormissement chez 31 étudiants après prise de caféine et prise d'un placebo.

- Résultats:

		Difficultés à l'endormissement		
		Oui	Non	
Boisson	Caféine	8	23	31
	Placebo	1	30	31
				62

Chi² $p = 0.34$

Test du χ^2 pour mesures répétées (Test de Mac Nemar)

- Analyse des résultats:
  - Difficultés à l'endormissement sous

Caféine et Placébo:

oui	oui	0
oui	non	8
non	oui	1
non	non	22
total:		31

Test du χ^2 pour mesures répétées (Test de Mac Nemar)

		Difficultés à l'endormissement sous caféine	
		Oui	Non
Difficultés à l'endormissement sous placebo	Oui	0	1
	Non	8	22

31

Test Mc Nemar: $p = 0.02$

VARIABLES DISCRETES	
2 groupes indépendants	
- comparaisons des proportions	Test de z
- comparaisons des nombres observés	Chi ^{2}
si valeur prédite < 5	Chi ^{2} avec correction de Yates
petits effectifs	Test exact de Fisher
Mesures répétées	
	Test de symétrie de McNemar
Plus de 2 groupes indépendants	
	Chi ^{2}

ANALYSIS OF VARIANCE

Pourquoi l'analyse de la variance (ANOVA)?

- Problème de multiplicité des tests
- Analyse des designs compliqués

COMPARAISON DE PLUS DE 2 MOYENNES

Nombres aléatoires distribués normalement avec une moyenne = 25, et SD = 5

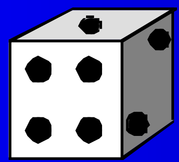
	1	2	3	4	5
	26.1	25.6	26.0	30.2	21.8
	26.1	29.3	27.1	33.1	20.2
	25.0	34.0	20.1	21.6	31.0
	18.9	18.5	20.7	24.1	29.6
	26.2	18.8	27.3	25.6	16.4
	25.7	27.8	25.6	34.2	31.4
	28.8	31.0	22.3	28.0	24.0
	21.3	20.6	25.5	33.3	19.3
	28.1	14.3	20.3	25.9	15.6
	26.1	20.5	20.5	32.8	20.9
Moyenne	25.2	24.0	23.5	28.9	23.0
SD	3.0	6.4	3.0	4.5	5.8
t-tests	"1-2"	"2-3"	"3-4"	"4-5"	
p-values	0.602	0.830	0.006	0.021	
	"1-3"	"2-4"	"3-5"		
	0.228	0.065	0.796		
	"1-4"	"2-5"			
	0.044	0.710			
	"1-6"				
	0.297				

Taux d'erreur α: 0.31

3 tests parmi les 10 comparaisons sont significatifs par chance uniquement

TAUX D'ERREUR DE TYPE I

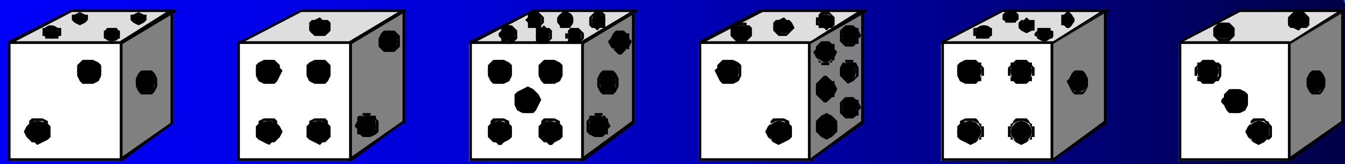
Taux d'erreur par comparaison: 0.17



$Pr_{(=1)} = 1/6 = 0.17$

$Pr_{(\neq 1)} = 1 - 0.17 = 0.83$

Taux d'erreur réel lié au design expérimental: 0.40

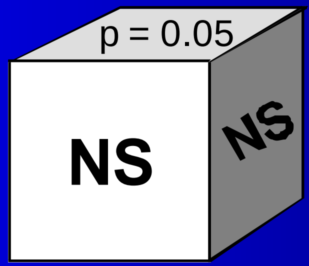


$Pr_{(=1)} = 6 * (1/6)^1 * (5/6)^5 = 0.40$

TAUX D'ERREUR DE TYPE I

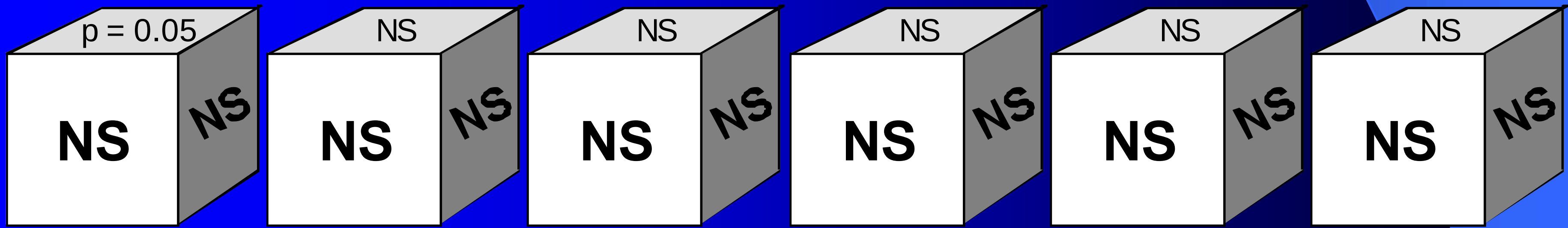
Taux d'erreur de type I par comparaison:

$\alpha = 0.05$



Taux d'erreur de type I lié au design expérimental:

$\alpha = 0.232$



$Pr = 6 * (0.05)^1 * (0.95)^5 = 0.232$

MULTIPLICITE

		Test t de Student	
Nombre de moyennes	Nombre de comparaisons $g = n * (n - 1) / 2$	Comparaison α	Expérimental a
2	1	0.05	0.05
3	3	0.05	0.1354
4	6	0.05	0.2321
5	10	0.05	0.3151
6	15	0.05	0.3658
7	21	0.05	0.3764

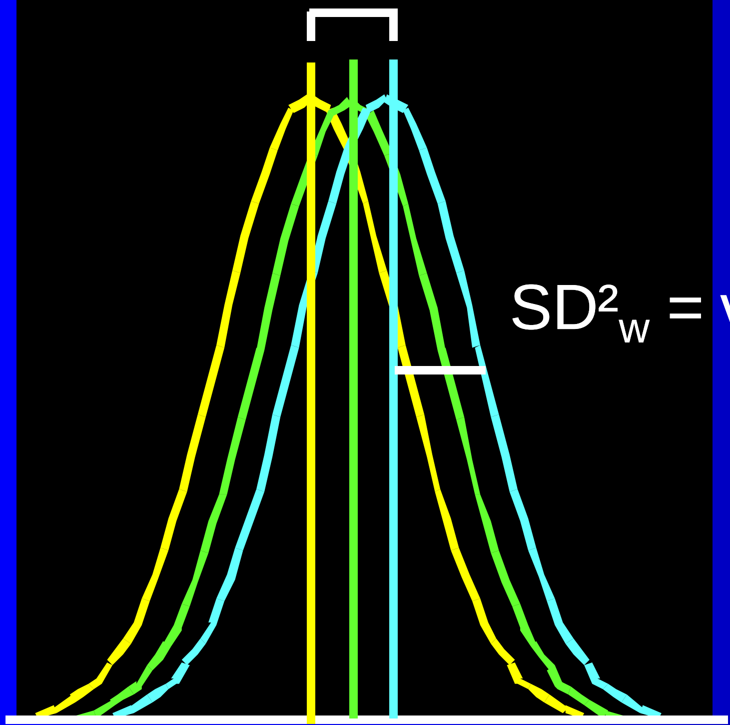
L'approximation la plus pessimiste (c-à-d., limite supérieure) du taux d'erreur expérimental = $1 - (1 - 0.05)^g$

Taux d'erreur expérimental = probabilité de faire au moins une erreur de type I pour *g* comparaisons

ANOVA

- L'analyse de la variance (ANOVA) compare plus de deux moyennes en utilisant un seul test global. Elle contrôle ainsi le taux d'erreur de type I expérimental au niveau de 5 % (α = 0.05)
- Si l'hypothèse nulle est vraie:

SD^2_b = variance entre groupes (between groups)

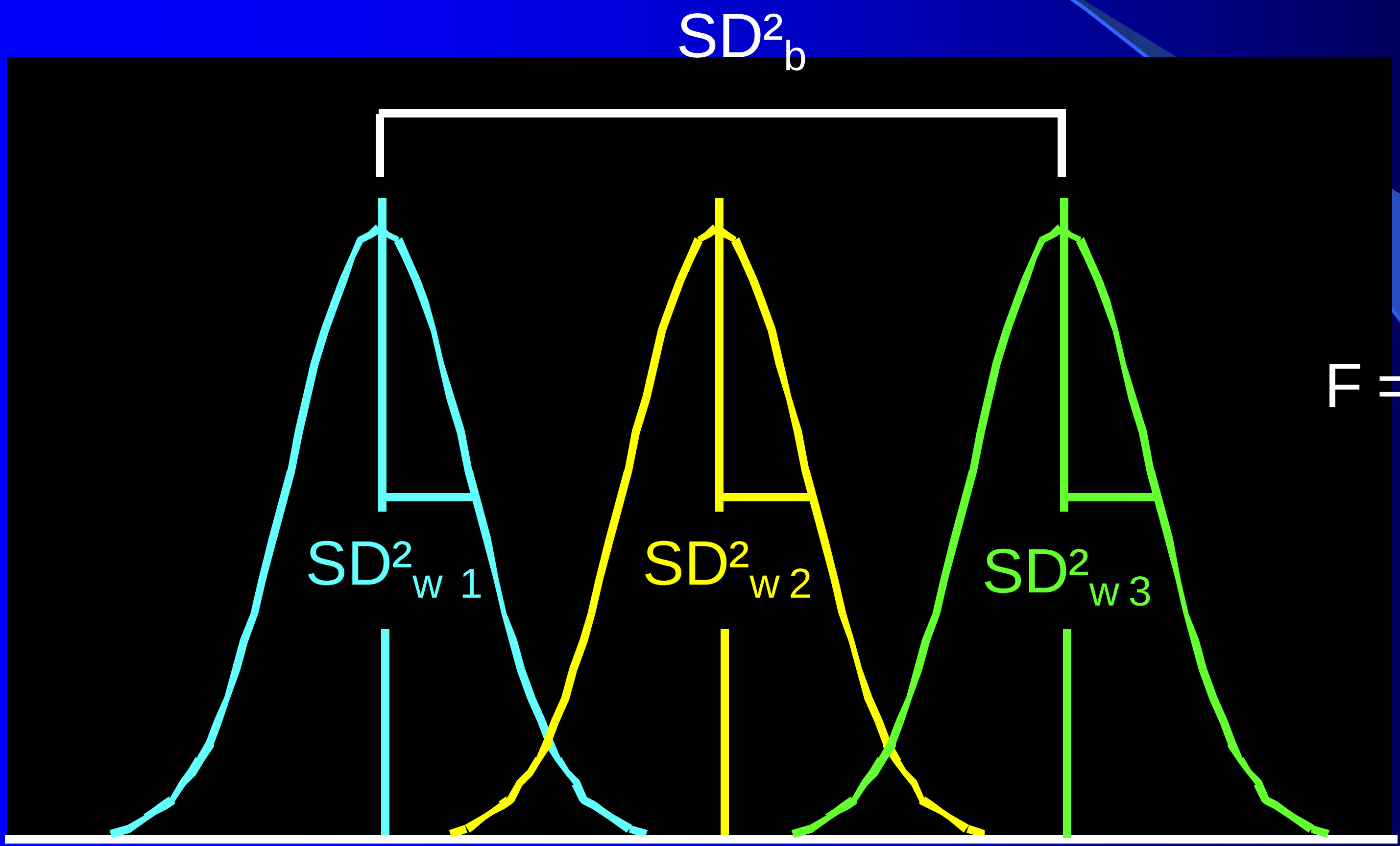


SD^2_w = variance intragroupe (within group)

$$F = \frac{SD^2_b}{SD^2_w} \leq 1 \text{ , } p = NS$$

ANOVA

- Si l'hypothèse nulle est rejetée:



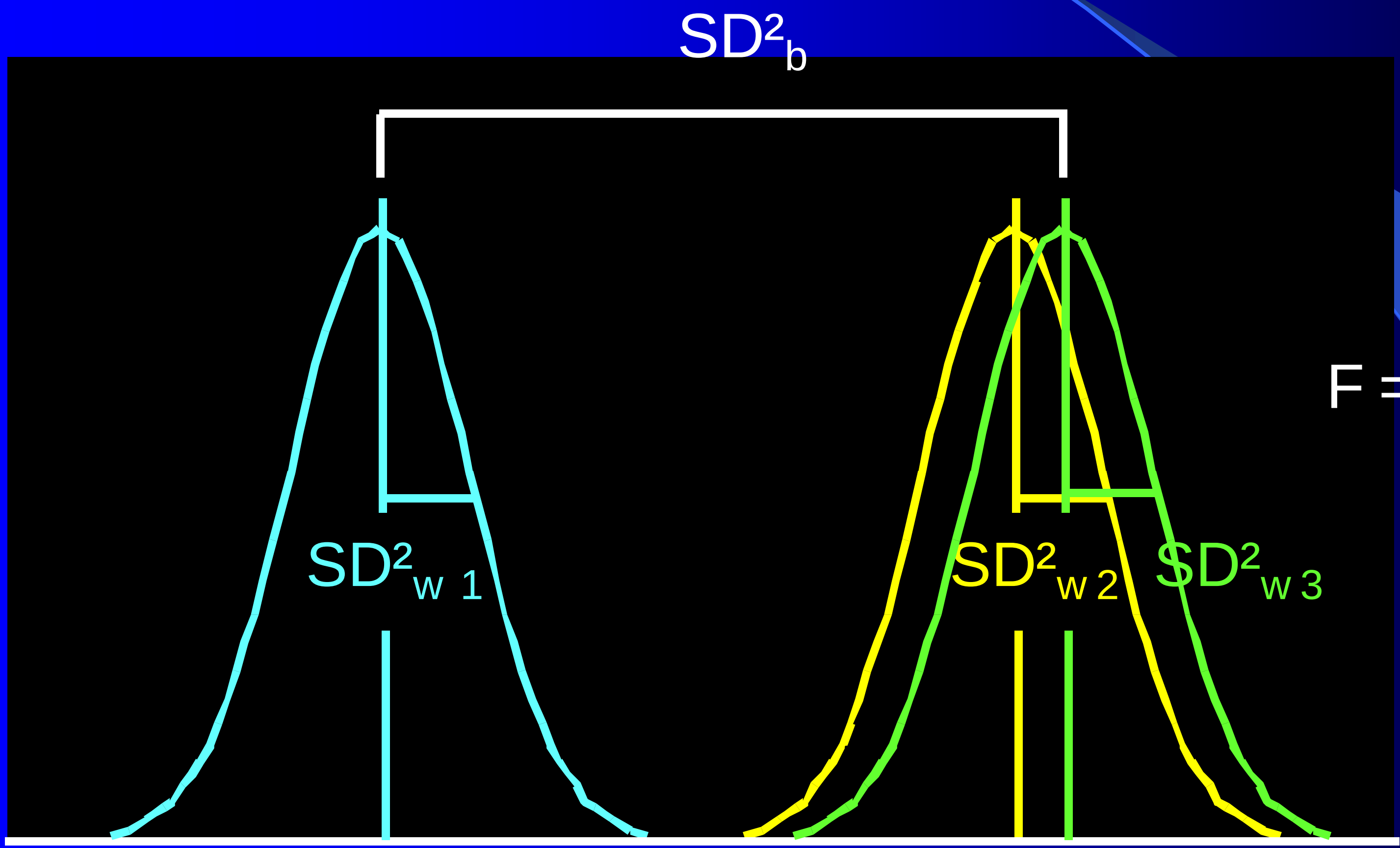
$$F = \frac{SD^2_b}{SD^2_w} > 1$$

$p < 0.05$

Condition: $SD^2_{w\,1} = SD^2_{w\,2} = SD^2_{w\,3} = SD^2_w$ (homoscédasticité)

ANOVA

- Si l'hypothèse nulle est rejetée:



$$F = \frac{SD^2_b}{SD^2_w} > 1$$

$p < 0.05$

Condition: $SD^2_{w\,1} = SD^2_{w\,2} = SD^2_{w\,3} = SD^2_w$ (homoscédasticité)

COMPARISON OF THREE MEANS (INDEPENDENT SAMPLES))

EXAMPLE:

Aim of the study:

efficacy of a β-blocking agent and a calcium channel blocker on systolic blood pressure.

Design:

15 hypertensive patients randomized in three groups of 5 patients: one group receiving a placebo, one group receiving a β-blocker, and one group receiving a calcium channel blocker.

COMPARISON OF THREE MEANS

(ANOVA ONE-WAY, PARAMETRIC APPROACH)

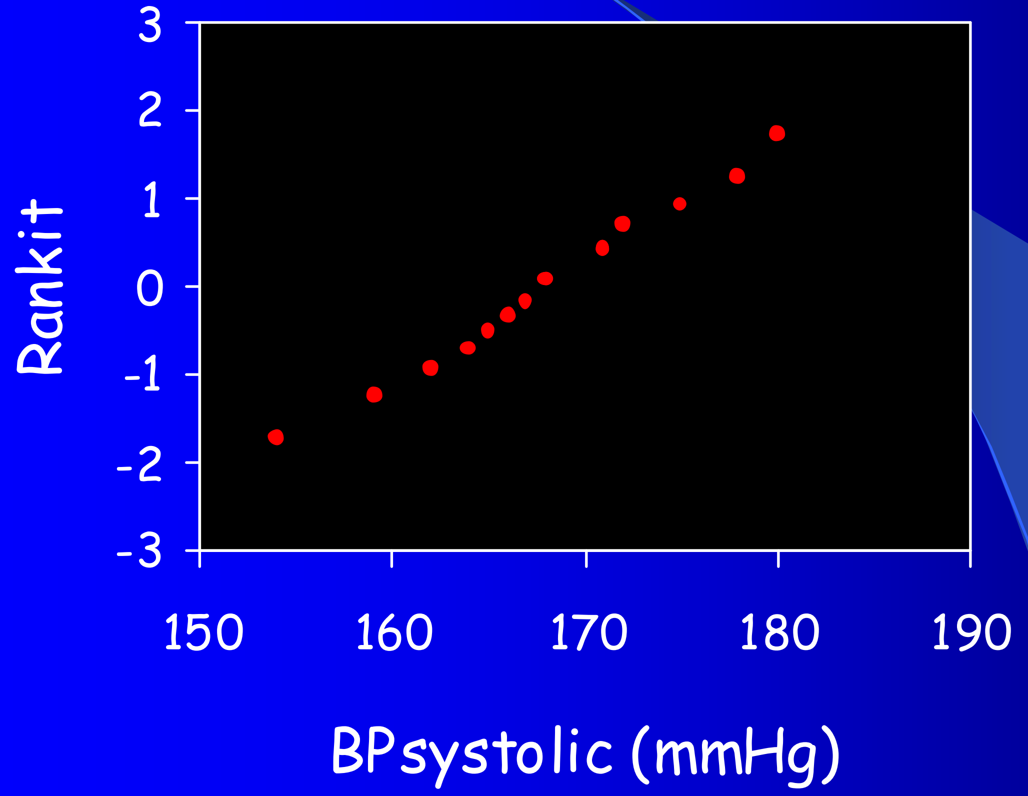
Results:

		Systolic blood pressure (mmHg)		
		Group 1 Placebo	Group 2 β blocker	Group 3 Calcium channel blocker
Patients	1	180	167	154
	2	175	171	166
	3	168	162	164
	4	171	168	159
	5	178	172	165
Mean		174.4	168.0	161.6
SD		4.9	3.9	5.0

COMPARISON OF THREE MEANS
(ANOVA ONE-WAY, PARAMETRIC APPROACH)

1) Check for the normality of the data

Rankit plot



Wilk-Shapiro = 0.9853

$$\text{Rankit} = z_{inv}\left(\frac{i - 0.375}{n + 0.25}\right)$$

COMPARISON OF THREE MEANS
(ANOVA ONE-WAY, PARAMETRIC APPROACH)

2) Check for the homoscedasticity of the variances

Group	Mean	Sample size	Group variance
1 2 3	174.4 168.0 161.6	5 5 5	24.3 15.5 25.3
Total	168.0	15	65.1

3 variances	Statistics	df	p
Bartlett's test of equal variance	$Chi^2 = 0.23$	1	0.6326
Hartley's test (highest var/lowest var) $\alpha < 0.05 \rightarrow F > 15.5$	$F = 1.632$	4	NS
Cochran's Q (highest var / Σ var) $\alpha < 0.05 \rightarrow Q > 0.7457$	$Q = 0.3886$	4	NS

COMPARISON OF THREE MEANS

(ANOVA ONE-WAY, PARAMETRIC APPROACH)

Group	Mean	Sample size	Group variance
1 2 3	174.4 168.0 161.6	5 5 5	24.3 15.5 25.3
Total	168.0	15	

$$MS = \frac{4 \times 24.3 + 4 \times 15.5 + 4 \times 25.3}{4 + 4 + 4} = 21.7$$

COMPARISON OF THREE MEANS

(ANOVA ONE-WAY, PARAMETRIC APPROACH)

3) Perform the analysis of variance

Group	Mean	Sample size	Group variance
1	174.4	5	24.3
2	168.0	5	15.5
3	161.6	5	25.3
Total	168.0	15	

ANOVA table

Source	df	SS	MS (variance)	F	p
Between groups	2	409.6	204.8	9.[illegible]4	0.[illegible]34
Within group	14-2= 12	670-409.6 = 260.4	21.7		
Total	14	670.0	47.9		

ANOVA ONE-WAY: COMPUTATIONS

Total Sum of Squares (Total SS)

		Systolic blood pressure (mmHg)		
		Group 1 Placebo	Group 2 β blocker	Group 3 Calcium channel blocker
Patients	1	180	167	154
	2	175	171	166
	3	168	162	164
	4	171	168	159
	5	178	172	165
Mean		174.4	168.0	161.6
SD		4.9	3.9	5.0

$\bar{y}_{..} = 168$

Total SS = $(180-168)^2 + (175-168)^2 + \ldots = 670$

Total df = $15 - 1 = 14$

ANOVA ONE-WAY: COMPUTATIONS

Treatment Sum of Squares (Treat SS)

		Systolic blood pressure (mmHg)		
		Group 1 Placebo	Group 2 β blocker	Group 3 Calcium channel blocker
Patients	1	174.4	168.0	161.6
	2	174.4	168.0	161.6
	3	174.4	168.0	161.6
	4	174.4	168.0	161.6
	5	174.4	168.0	161.6
Mean		174.4	168.0	161.6
SD		0	0	0

$\bar{y}_{..} = 168$

Treat SS = $5 * (174.4-168)^2 + 5 * (168.0-168)^2 + 5 * (161.6-168)^2 = 409.6$

Treat df = $3 - 1 = 2$

STRUCTURE OF THE VARIANCE

Sum of squares (SS)

Mean square (MS)

$$Var(x) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n - 1}$$

Degree of freedom (df)

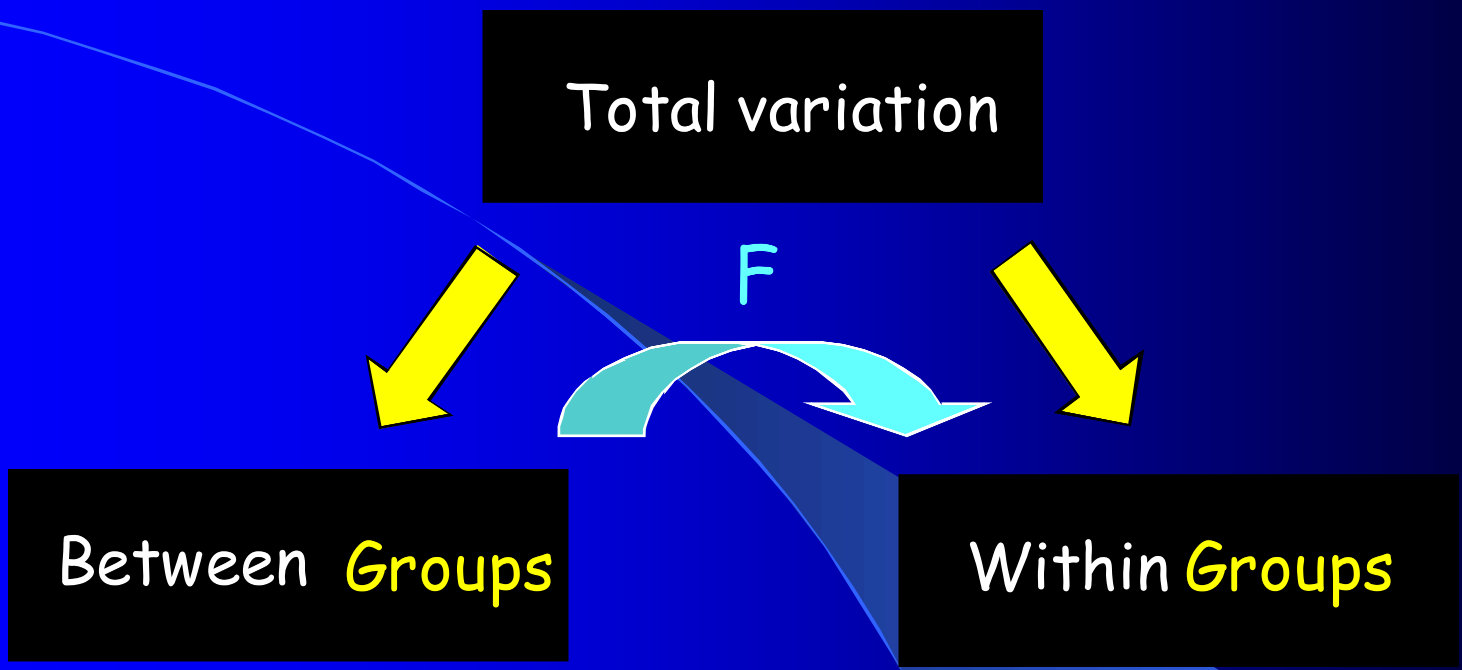
COMPARISON OF THREE MEANS (ANOVA ONE-WAY)

ANOVA table:

Source of variation	Degree of freedom (df)	Sum of Squares (SS)	Mean square (MS = SS / df)	F (MSb / MSw)	p
Between groups	2	409.6	204.8	9.44	0.0034
Within group (error or residual)	14-2= 12	670-409.6 = 260.4	21.7		
Total	14	670.0	47.9		

COMPARISON OF THREE MEANS

(ANOVA ONE-WAY)



If F significant ($p < 0.05$), pairwise comparisons must be performed.

If F non significant ($p \geq 0.05$), no further analysis is needed.

PAIRWISE COMPARISONS OF MEANS AFTER A SIGNIFICANT ANOVA

Modified Student 's t test (Fisher's protected t test)

$$t_{(df\ error)} = \frac{|\ m_1 - m_2\ |}{\sqrt{\dfrac{MSerror}{n_1} + \dfrac{MSerror}{n_2}}} \quad \text{with } df_{error}$$

Least Significant Difference (LSD)

$$LSD = t_{(df\ error,\ 0.05)} * \sqrt{\frac{MSerror}{n_1} + \frac{MSerror}{n_2}}$$

LSD does not protect against the type I experimentwise error rate

- Bonferroni's correction
- Sidak's correction
- Holm's procedure
- Scheffé's test
- Tukey's test
- Dunnett's test (control group)
- Duncan's test
- Student-Newman-Keuls test

COMPARISON OF SEVERAL MEANS

Random numbers normally distributed with mean = 25, and SD = 5

	1	2	3	4	5
	26.1	25.6	26.0	30.2	21.8
	26.1	29.3	27.1	33.1	20.2
	25.0	34.0	20.1	21.6	31.0
	18.9	18.5	20.7	24.1	29.6
	26.2	18.8	27.3	25.6	16.4
	25.7	27.8	25.6	34.2	31.4
	28.8	31.0	22.3	28.0	24.0
	21.3	20.6	25.5	33.3	19.3
	28.1	14.3	20.3	25.9	15.6
	26.1	20.5	20.5	32.8	20.9
Mean	25.2	24.0	23.5	28.9	23.0
SD	3.0	6.4	3.0	4.5	5.8
t-tests	"1-2"	"2-3"	"3-4"	"4-5"	
p-values	0.602	0.830	0.006	0.021	
	"1-3"	"2-4"	"3-5"		
	0.228	0.065	0.796		
	"1-4"	"2-5"			
	0.044	0.710			
	"1-6"				
	0.297				

α error rate: 0.31

Spurious significant t-tests occurred in 3 tests among 10 comparisons

THE BONFERRONI's
CORRECTION OF THE p-VALUE

- Either, correction of the level of significance:

  $a = \alpha$ / number of comparisons
  If 3 comparisons, 0.05 / 3 = 0.0166

  Therefore, we declare significant all $p < 0.0166$

- Either, correction of the p-value:

  Corrected p-value = p-value * number of comparisons

  For 3 comparisons, a p-value = 0.0166 becomes
  corrected p-value = 0.0166 * 3 = 0.05

COMPARISON OF SEVERAL MEANS

Bonferroni 's correction: p-value * number of comparisons (10)

	1	2	3	4	5
	26.1	25.6	26.0	30.2	21.8
	26.1	29.3	27.1	33.1	20.2
	25.0	34.0	20.1	21.6	31.0
	18.9	18.5	20.7	24.1	29.6
	26.2	18.8	27.3	25.6	16.4
	25.7	27.8	25.6	34.2	31.4
	28.8	31.0	22.3	28.0	24.0
	21.3	20.6	25.5	33.3	19.3
	28.1	14.3	20.3	25.9	15.6
	26.1	20.5	20.5	32.8	20.9
Mean	25.2	24.0	23.5	28.9	23.0
SD	3.0	6.4	3.0	4.5	5.8
t-tests	"1-2"	"2-3"	"3-4"	"4-5"	
corrected p-values	1.000	1.000	0.058	0.205	
	"1-3"	"2-4"	"3-5"		
	1.000	0.651	1.000		
	"1-4"	"2-5"			
	0.443	1.000			
	"1-6"				
	1.000				

True α error rate: = 0.047, with the Bonferroni's correction

No significant t-test occurred among 10 comparisons

MULTIPLICITY

		Student's t test		Bonferroni's correction		Sidak's correction	
Number of means	Number of comparisons $g = n * (n - 1) / 2$	Comparison α	Experiment a	Comparison $p = \alpha / g$	Experiment a	Comparison $p = 1 - (1-\alpha)^{1/g}$	Experiment a
2	1	0.05	0.05	0.05	0.05	0.05	0.05
3	3	0.05	0.1354	0.0167	0.0492	0.0170	0.05
4	6	0.05	0.2321	0.0083	0.0490	0.0085	0.05
5	10	0.05	0.3151	0.005	0.0489	0.0051	0.05
6	15	0.05	0.3658	0.0033	0.0489	0.0034	0.05
7	21	0.05	0.3764	0.0024	0.0488	0.0024	0.05

Pessimistic approximation (i.e., upper bound) of experimentwise error rate = $1 - (1 - 0.05)^g$

Experimentwise error rate = probability of making at least 1 type I error for g comparisons

HOLM's SEQUENTIALLY REJECTIVE PROCEDURE

- Holm (1979) proposed a procedure which is more powerful than single step Bonferroni's or Sidak's procedures and that controls experimentwise error rate

- Briefly, the m p-values are ordered (i from 1 to m) in an increasing order and the level of significance is given by:

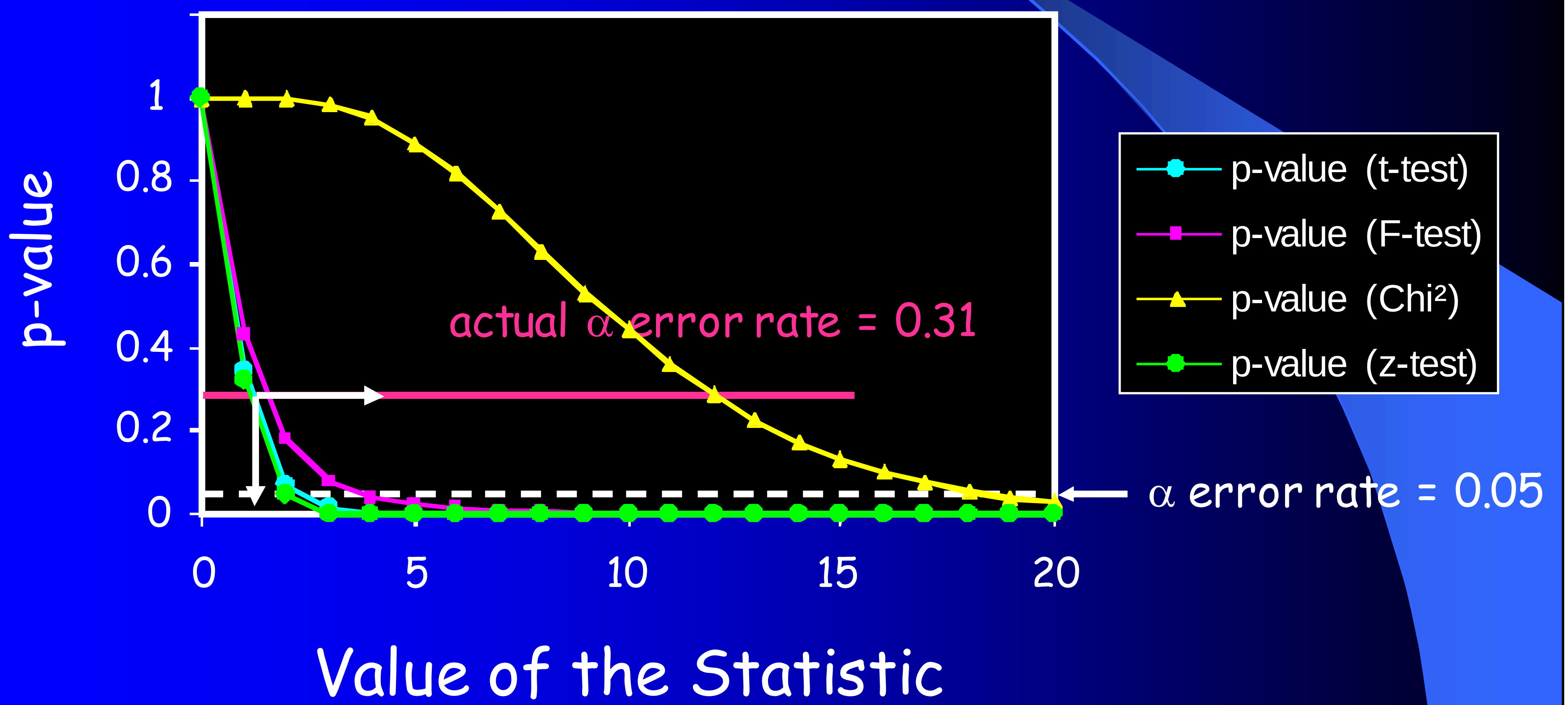
$$p_i \leq \frac{\alpha}{(m-i+1)}$$

HOLM's SEQUENTIALLY REJECTIVE PROCEDURE

m1 = 174.4	m2 = 168.0	m3 = 161.6

Ordred treatment differences	Order (i)	Ordered p-values (modified t-tests)	Holm's p-value P = 0.05/ (m-i+1) with m = 3	Significance of the Holm's procedure
m1 - m3 = 12.8	1	0.00095386	0.0166666	$p < 0.05$
m1 - m2 = 6.4	2	0.05058615	0.025	p NS
m2 - m3 = 6.4	3	0.05058615	0.05	p NS

RELATIONSHIP BETWEEN THE p-VALUE AND THE VALUE OF THE STATISTIC



PAIRWISE COMPARISON IN THE SITUATION OF MULTIPLE COMPARISONS

- Correction of the level of significance (α) of the classical t statistics:
  - Bonferroni's correction
  - Sidak's correction

- Changing the value of the statistics:
  - Scheffé's test
  - Tukey & Tukey-Kramer's test
  - Dunnett's test (control)

- Multiple levels of significance depending on the ranking of the difference:
  - Holm's procedure
  - Student Newman Keuls
  - Duncan's test

Tests post-hoc (1)

- Après avoir mis en évidence un test F significatif, il est nécessaire de déterminer quelle(s) paire(s) de moyennes sont statistiquement significatives.
- Ces comparaisons multiples posent problème vu la dérive de l'erreur de type I.
- Une approche est de déterminer les différences significatives entre les moyennes en utilisant des test de t multiples spéciaux. Il s'agit des procédures de comparaisons multiples ou comparaisons post-hoc.

Tests post-hoc (2)

- Les tests post-hoc ne sont pas disponibles pour les designs avec mesures répétées (facteurs intra-sujets). Ceci est lié au fait que les tests ont été développés sur l'hypothèse de l'indépendance des différentes moyennes comparées. Cette assomption est violée par la nature même d'un design avec mesures répétées (mesures corrélées).
- Cependant il est possible de réaliser des contrastes impliquant des facteurs intra-sujets, à la condition que ces contrastes soient précisés avant la réalisation du protocole.
- A l'exception du test de Dunnett, les tests post-hoc sont tous des tests bilatéraux.

Post-hoc tests (3)

- Tous les tests post-hoc doivent remplir une ou plusieurs des conditions suivantes:
  - Homogénéité de la variance, c-à-d, variance égale entres cellules.
  - Comparabilité de la taille des cellules, c-à-d, n égal de cellules en cellules
  - Normalité de la distribution des données dans chaque cellule.
  - Un test de F significatif (omnibus F-test).

Post-hoc tests (4)

Procedure	Associated hypothesis	Preceded by a significant F?	Homogeneity by cell	Equal cell n	Cell normality	Max number of means
LSD (F protected)	null	yes	yes	yes	yes	no limit
Tukey	null	no	yes	yes	yes	20
Tukey-Kramer	null	no	yes	no	yes	20
SNK	null	yes	yes	yes	yes	20
Duncan	null	no	yes	no	yes	20
Scheffé	null	yes	no	no	no	no limit
Bonferroni	null	no	yes	yes	yes	no limit
Dunnett	null or directional	no	no	no	yes	20 two-tail & 9 one-tail
Holm	null	no	yes	yes	no	no limit
Sidak	null	no	yes	yes	yes	no limit
Games-Howell	null	yes	no	no ($\geq$ 6)	yes	20

Post-hoc tests (5)

Procedure	Error summary		
	Comparisonwise		Experimentwise
LSD (F protected) Modified t-test	$p = \alpha$ per comparison	and	$p > \alpha$ per set of comparisons
Tukey	$p \leq \alpha$ per comparison	and	$p = \alpha$ per set of comparisons
Tukey-Kramer	$p \leq \alpha$ per comparison	and	$p = \alpha$ per set of comparisons
SNK	$p = \alpha$ by layer of comparison	and	$p > \alpha$ per set of comparisons
Duncan	$p = \alpha$ for comparison of only 2 means	but	$p > \alpha$ as number of means increases
Scheffé	$p < \alpha$ per comparison	and	$p < \alpha$ per set of comparisons
Bonferroni	$p < \alpha$ per comparison	and	$p \leq \alpha$ per set of comparisons
Dunnett	$p < \alpha$ per set of comparisons		
Holm	$p < \alpha$ per comparison	and	$p = \alpha$ per set of comparisons
Sidak	$p < \alpha$ per comparison	and	$p = \alpha$ per set of comparisons
Games-Howell	$p = \alpha$ per comparison	and	$p = \alpha$ per set of comparisons

COMPARISON OF THREE MEANS

(ANOVA ONE-WAY, PARAMETRIC APPROACH)

Results:

		Systolic blood pressure (mmHg)		
		Group 1 Placebo	Group 2 β blocker	Group 3 Calcium channel blocker
Patients	1	180	167	154
	2	175	171	166
	3	168	162	164
	4	171	168	159
	5	178	172	165
Mean SD		174.4 4.9	168.0 3.9	161.6 5.0

COMPARISON OF THREE MEANS
(ANOVA ONE-WAY, PARAMETRIC APPROACH)

4) Pairwise comparisons

Procedure	Critical value	Minimum significant difference (MSD)	$1 \rightarrow 2$ $\Delta = 6.4$	$1 \rightarrow 3$ $\Delta = 12.8$	$2 \rightarrow 3$ $\Delta = 6.4$
LSD	2.179	6.419	NS	S	NS
Bonferroni	2.779	8.188	NS	S	NS
Sidak	2.770	8.161	NS	S	NS
Tukey (HSD)	3.783	7.881	NS	S	NS
Scheffé	3.885	8.212	NS	S	NS
Holm	2.779 - 2.179	8.188 – 6.419	NS	S	NS
Dunnett	2.11	6.216	S	S	-
Duncan	3.08-3.23	6.416 – 6.729	NS	S	NS
SNK	3.08-3.77	6.416 – 7.854	NS	S	NS

Contrastes orthogonaux

- Permettent la comparaison de combinaisons linéaires des moyennes.
- La somme des coefficients doit être égale à zéro
- Exemple: comparaison du groupe contrôle avec les deux groupes traités:
  - Coefficient: 2 – 1 - 1 ($\Sigma = 0$)
  - Ho: $2*\text{moyenne}_1 - \text{moyenne}_2 - \text{moyenne}_3 = 0$

    $2*\text{moyenne}_1 = \text{moyenne}_2 + \text{moyenne}_3$
  - Contraste: 19.20 (SE = 5.103) $\rightarrow$ t = 3.76 $\rightarrow$ p = 0.0027

Contrastes polynomiaux

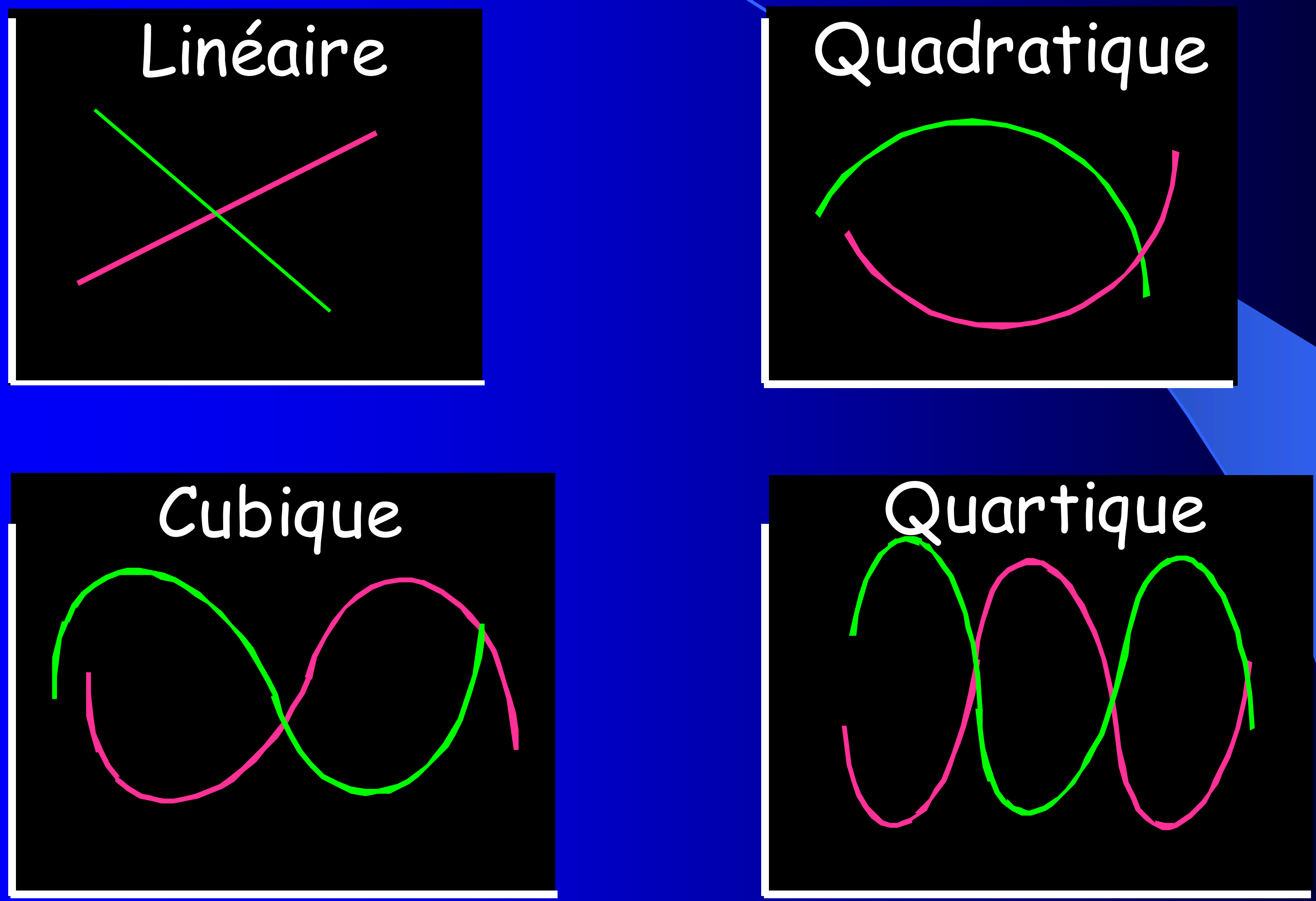
- Permettent de mettre en évidence une relation linéaire (x), quadratique (x^2), cubique (x^3), quartique (x^4) entre les moyennes.

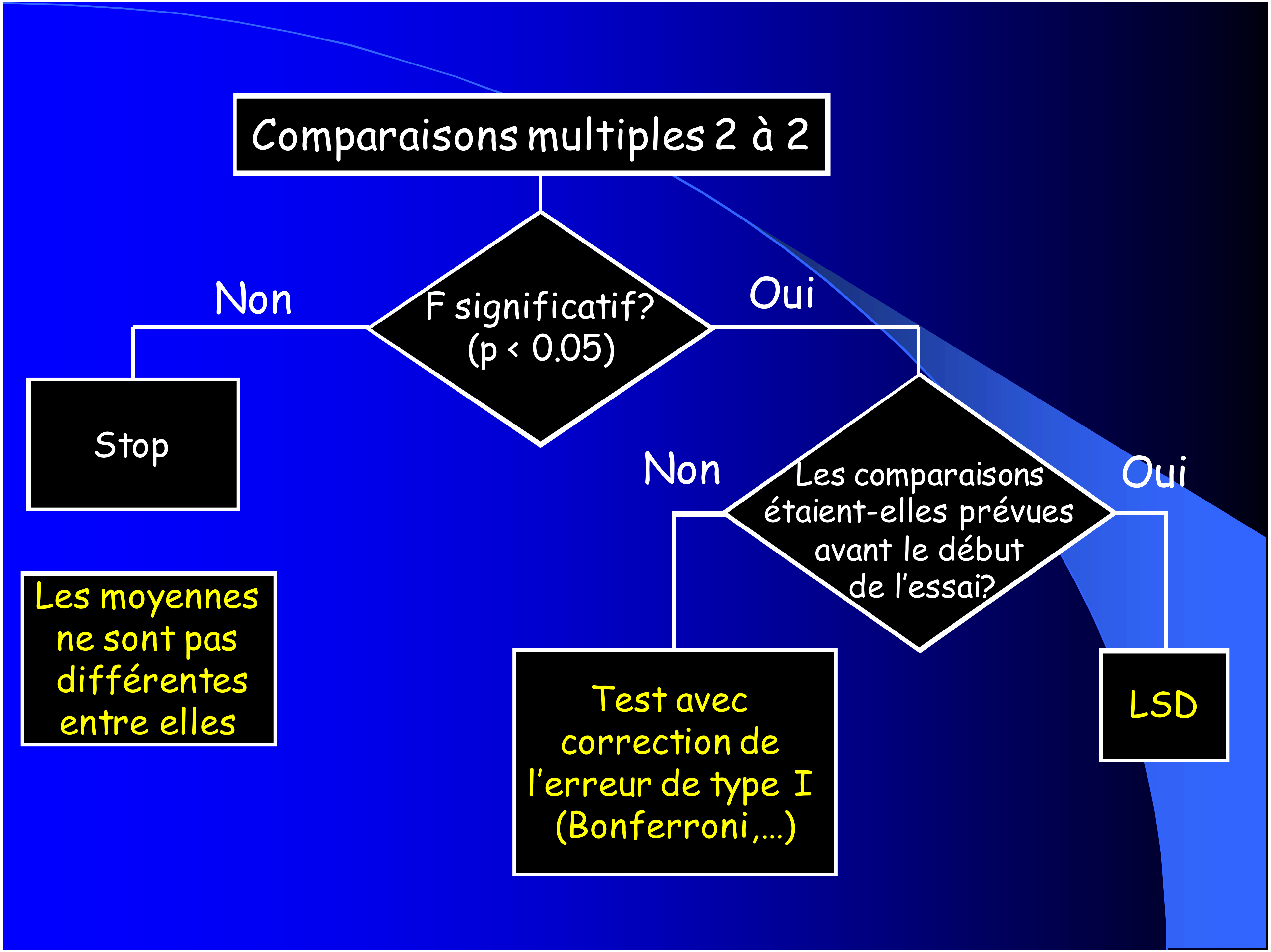
- La somme des coefficients doit être égale à zéro

- Exemple:

Degré du polynome	SS	F	p
1 (linéaire)	406.6	18.88	0.001
2 (quadratique)	0.0001	0.00	1.000

Contrastes polynomiaux




Comparaisons multiples 2 à 2
Non
F significatif?
(p < 0.05)
Oui
Stop
Les moyennes ne sont pas différentes entre elles
Non
Les comparaisons étaient-elles prévues avant le début de l'essai?
Oui
Test avec correction de l'erreur de type I (Bonferroni,…)
LSD

COMPARISON OF THREE GROUPS: Kruskal-Wallis (non parametric approach)

Results:

		Systolic blood pressure (mmHg)		
		Group 1 Placebo	Group 2 β blocker	Group 3 Calcium channel blocker
Patients	1	180	167	154
	2	175	171	166
	3	168	162	164
	4	171	168	159
	5	178	172	165
Mean		174.4	168.0	161.6
SD		4.9	3.9	5.0

COMPARISON OF THREE GROUPS: Kruskal-Wallis (non parametric approach)

1) Rank the values:

Ordered values	Rank	Ties
154	1	
159	2	
162	3	
164	4	
165	5	
166	6	
167	7	
168	8	8.5
168	9	8.5
171	10	10.5
171	11	10.5
172	12	
175	13	
178	14	
180	15	

COMPARISON OF THREE GROUPS: Kruskal-Wallis (non parametric approach)

2) Replace the values by their rank:

		Systolic blood pressure (mmHg)		
		Group 1 Placebo	Group 2 β blocker	Group 3 Calcium channel blocker
Patients	1	15	7	1
	2	13	10.5	6
	3	8.5	3	4
	4	10.5	8.5	2
	5	14	12	5
Mean rank		12.2	8.2	3.6

COMPARISON OF THREE GROUPS: Kruskal-Wallis (non parametric approach)

3) Compute ANOVA with the ranks:

ANOVA table

Source	df	SS	MS (variance)	F	p
Between subjects	2	185.2	92.6	11.85	0.0014
Within subjects	14-2= 12	279-185.2 = 93.8	7.817		
Total	14	279.0			

Kruskal-Wallis statistic: 9.2932
p-value = 0.0096

COMPARISON OF THREE GROUPS: Kruskal-Wallis (non parametric approach)

4) Compute pairwise comparisons using U tests

Comparisons	U stat	p (1 tail)	p (2 tail)	Bonferroni's correction
Group 1 & Group 2	21 vs 4	0.046	0.0947	0.284
Group 1 & Group 3	25 vs 0	0.004	0.0122	0.0366
Group 2 & Group 3	22 vs 3	0.0278	0.0601	0.1803

REPEATED MEASURES

- Problem with correlation between measurements on the same subject.
- Correlation is calculated from variance and covariance
- How can we take into account of correlation?

COMPARISON OF THREE MEANS (REPEATED MEASURES)

EXAMPLE:

Aim of the study:

efficacy of a β-blocking agent and a calcium channel blocker on systolic blood pressure.

Design:

Repeated measures in 5 hypertensive patients: time 1: after a placebo, time 2: after a β-blocker, and time 3: after a calcium channel blocker.

COMPARISON OF THREE MEANS
(ANOVA ONE-WAY FOR REPEATED MEASURES)

Results:

		Systolic blood pressure (mmHg)				
		Time 1 Placebo	Time 2 β blocker	Time 3 Calcium channel blocker	Mean	SD
Patients	1	180	167	154	167.0	13.0
	2	175	171	166	170.7	4.5
	3	168	162	164	164.7	3.1
	4	171	168	159	166.0	6.2
	5	178	172	165	171.7	6.5
Mean SD		174.4 4.9	168.0 3.9	161.6 5.0	$\bar{y}_{..} = 168$	

Total SS = $(180-168)^2 + (175-168)^2 + \ldots = 670$

Total df = 15 – 1 = 14

COMPARISON OF THREE MEANS
(ANOVA ONE-WAY FOR REPEATED MEASURES)

Treatment Sum of Squares (Treat SS)

		Systolic blood pressure (mmHg)				
		Time 1 Placebo	Time 2 β blocker	Time 3 Calcium channel blocker	Mean	SD
Patients	1	174.4	168.0	161.6	-	-
	2	174.4	168.0	161.6	-	-
	3	174.4	168.0	161.6	-	-
	4	174.4	168.0	161.6	-	-
	5	174.4	168.0	161.6	-	-
Mean SD		174.4 0	168.0 0	161.6 0	$\bar{y}_{..} = 168$	

Treat SS = 5 * $(174.4-168)^2$ + 5 * $(168.0-168)^2$ + 5 * $(161.6-168)^2$ = 409.6

Treat df = 3 – 1 = 2

COMPARISON OF THREE MEANS
(ANOVA ONE-WAY FOR REPEATED MEASURES)

Between Subjects Sum of Squares (Between SS)

		Systolic blood pressure (mmHg)				
		Time 1 Placebo	Time 2 β blocker	Time 3 Calcium channel blocker	Mean	SD
Patients	1	167.0	167.0	167.0	167.0	0
	2	170.7	170.7	170.7	170.7	0
	3	164.7	164.7	164.7	164.7	0
	4	166.0	166.0	166.0	166.0	0
	5	171.7	171.7	171.7	171.7	0
Mean		-	-	-	$\bar{y}_{..} = 168$	
SD		-	-	-		

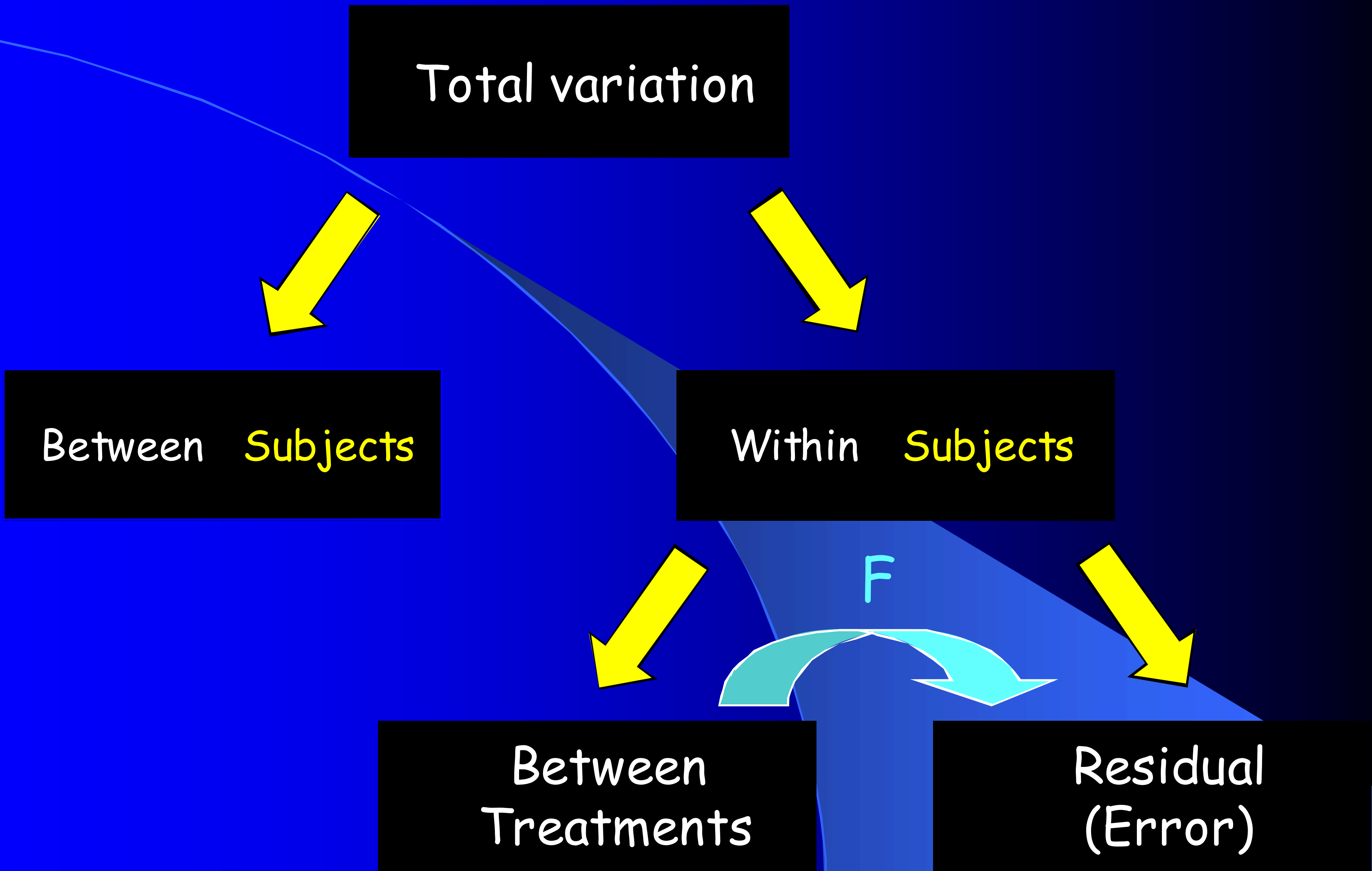
Between SS = 3 * $(167.0-168)^2$ + 3 * $(170.7-168)^2$ + … = 110.0
Between df = 5 – 1 = 4

COMPARISON OF THREE MEANS
(ANOVA ONE-WAY FOR REPEATED MEASURES)

ANOVA table

Source of variation	Degree of freedom (df)	Sum of Squares (SS)	Mean square (MS = SS / df)	F (MSb / MSw)	p
Between subjects	4	110.0	27.5	1.46	0.299
Within subjects - treatments - error or residual	 2 8	 409.6 150.4	 204.8 18.8	 10.9	 0.005
Total	14	670.0	47.9		

COMPARISON OF THREE MEANS
(ANOVA ONE-WAY FOR REPEATED MEASURES)



COMPARISON OF THREE MEANS:
Friedman (non parametric approach)

Results:

		Systolic blood pressure (mmHg)				
		Time 1 Placebo	Time 2 β blocker	Time 3 Calcium channel blocker	Mean	SD
Patients	1	180	167	154	167.0	13.0
	2	175	171	166	170.7	4.5
	3	168	162	164	164.7	3.1
	4	171	168	159	166.0	6.2
	5	178	172	165	171.7	6.5
Mean SD		174.4 4.9	168.0 3.9	161.6 5.0	$\bar{y}_{..} = 168$	

COMPARISON OF THREE MEANS:

Friedman (non parametric approach)

1) Rank the values:

Ordered values	Rank	Ties
154	1	
159	2	
162	3	
164	4	
165	5	
166	6	
167	7	
168	8	8.5
168	9	8.5
171	10	10.5
171	11	10.5
172	12	
175	13	
178	14	
180	15	

COMPARISON OF THREE MEANS:
Friedman (non parametric approach)

2) Replace the values by their rank:

		Systolic blood pressure (mmHg)				
		Time 1 Placebo	Time 2 β blocker	Time 3 Calcium channel blocker	Mean rank	
Patients	1	15	7	1	7.6	
	2	13	10.5	6	9.8	
	3	8.5	3	4	5.16	
	4	10.5	8.5	2	7.0	
	5	14	12	5	10.3	
Mean rank		12.2	8.2	3.6		

COMPARISON OF THREE MEANS:
Friedman (non parametric approach)

3) Compute ANOVA with the ranks:

ANOVA table

Source of variation	Degree of freedom (df)	Sum of Squares (SS)	Mean square (MS = SS / df)	F	p
Between subjects	4	53.833	13.458	2.69	0.108
Within subjects - treatments - error or residual	 2 8	 185.2 39.967	 92.6 4.9958	 18.54	 0.001
Total	14	279.0			

Friedman statistic:

- treatments 8.40 p = 0.015
- between subjects 6.133 p = 0.189

COMPARISON OF THREE MEANS:
Friedman (non parametric approach)

4) Compute pairwise comparisons using Wilcoxon signed rank test

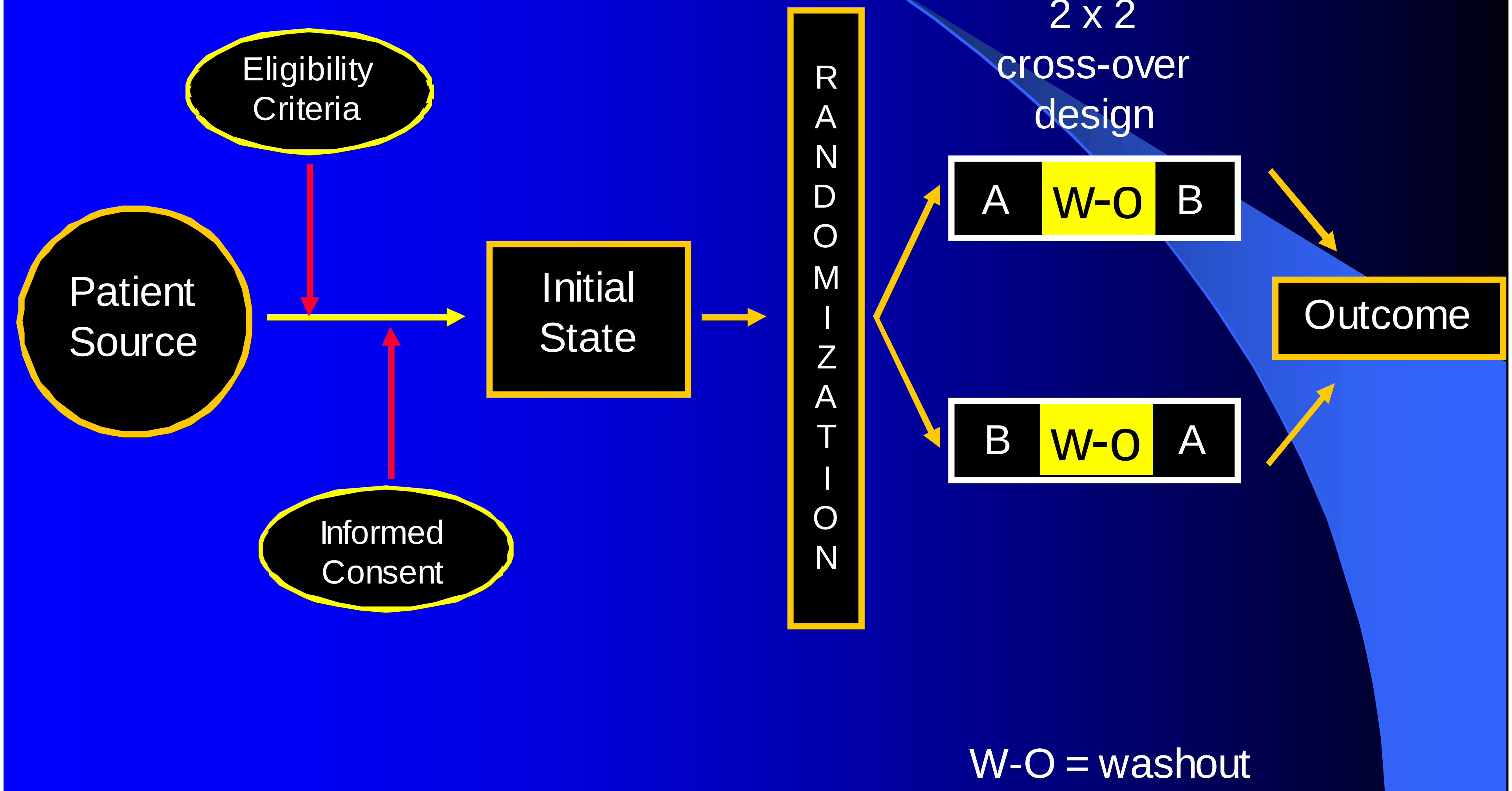
Comparisons	Sum negative & positive ranks	p (1 tail)	p (2 tail)
Mean 1 & Mean 2	0 vs 15	0.0312	0.0591
Mean 1 & Mean 3	0 vs 15	0.0312	0.0591
Mean 2 & Mean 3	-1 vs 14	0.0625	0.1056

VARIABLES CONTINUES	Distribution gaussienne Tests paramétriques	Distribution non-gaussienne Tests non paramétriques
2 moyennes (groupes indépendants)		
- variances égales	Test t non pairé	Mann-Withney U test Median test
- variances inégales	Test t non pairé (modification de Satterthwaite)	
2 moyennes (groupes pairés)		
	Test t pairé	Sign test Wilcoxon signed rank test
Plus de 2 moyennes (groupes indépendants)		
- variances intra-groupes égales	ANOVA à 1 facteur	Kruskal-Wallis
- variances intra-groupes inégales	→	
Plus de 2 moyennes (mesures répétées)		
	ANOVA à 2 facteurs	Friedman

Why do we need ANOVA?

- Multiplicity Problem
- Complicated design analysis

RCT with cross-over design



2 x 2 CROSS-OVER DESIGN: an example

	Nicardipine	Placebo
Group 1	16 26 8 37 9 41 52 10 11 30	12 19 20 44 25 36 36 11 20 27
Mean SD	24.0 15.6	25.0 10.8
	Placebo	Nicardipine
Group 2	18 12 46 51 28 29 51 46 18 44	12 4 37 58 2 18 44 14 30 4
Mean SD	34.3 15.0	22.3 19.1

- Randomized double-blind cross-over trial comparing nicardipine and placebo in patients with Raynaud's phenomenon.
- The data are the number of attacks (Raynaud's phenomenon) in two weeks. There was a one-week wash-out period between the two treatments periods.

Kahan A. et al. Angiology 1987;38:333-337.

2 x 2 CROSS-OVER DESIGN

- Estimation of the effects:

	PERIOD 1	PERIOD 2	Total	Differences
GROUP 1 (sequence 1)	A1	B1	A1 + B1	D1 = A1 – B1
GROUP 2 (sequence 2)	B2	A2	B2 + A2	D2 = B2 – A2
Total	A1 + B2	B1 + A2		

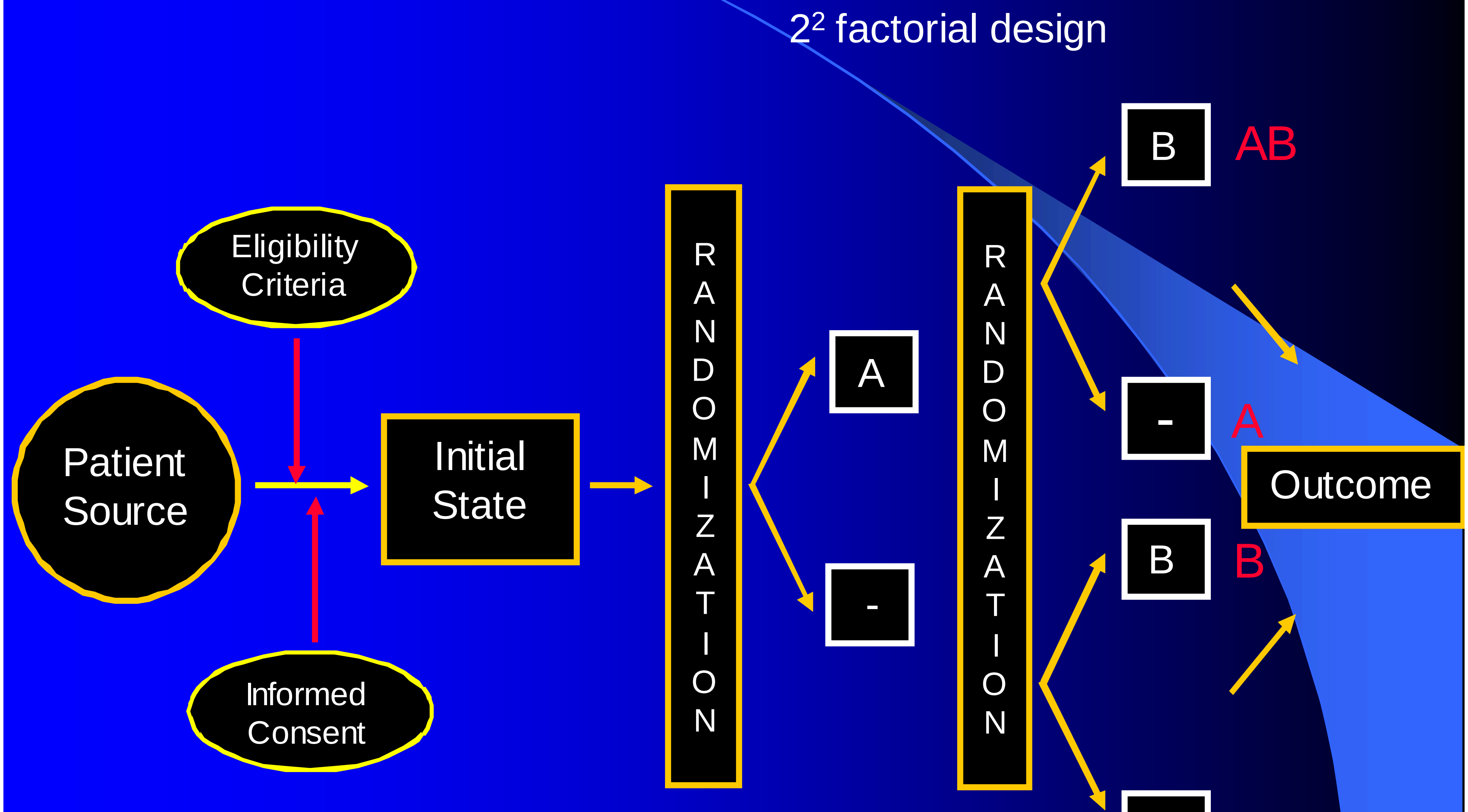
- ✓ Carry-over effect (group effect):
  (A1 + B1) vs (B2 + A2)
- ✓ Time effect (period effect):
  (A1 + B2) vs (B1 + A2) = (A1 - B1) vs - (B2 - A2)
- ✓ Treatment effect (interaction period x group):
  (A1 + A2) vs (B1 + B2) = (A1 - B1) vs (B2 - A2)

ANOVA: CROSS-OVER DESIGN

Model: group group*subj(E1) period group*period group*period*subj (E2)

SOURCE OF VARIATION	SUM OF SQUARES	df	MEAN SQUARE	F	p
Between groups					
GROUP (CARRY OVER effect	144.4	1	144.4	0.37	0.548
ERROR 1 (GROUP X SUBJECT)	6928.2	18	384.9		
Within groups					
PERIOD (TIME effect)	302.5	1	302.5	3.32	0.085
GROUP x PERIOD (TREATMENT effect)	422.5	1	422.5	4.64	0.045
ERROR 2 (GROUP x PERIOD x SUBJECT)	1640.0	18	91.1		
Total	9437.6	39			

RCT with 2^2 factorial design



A factor is an assigned variable: drug, diet, …

2^2 FACTORIAL DESIGN: an example

Cholesterol level (mg/100 ml)		DRUG	
		No	Yes
DIET	No	X 255 287 292 278 298	Xb 221 231 198 195 185
	Yes	Xa 245 235 240 252 260	Xab 165 185 201 198 182

- Randomized double-blind factorial design comparing a diet and a drug in hypercholesterolemic patients.
- The data are the level of cholesterol in mg/100ml.

2^2 FACTORIAL DESIGN

- Estimation of the effects:

		TREATMENT B	
		No	Yes
TREATMENT A	No	X	Xb
	Yes	Xa	Xab

✓ Effect of treatment A:
$\frac{1}{2}\{(Xa - X) + (Xab - Xb)\}$

✓ Effect of treatment B:
$\frac{1}{2}\{(Xb - X) + (Xab - Xa)\}$

✓ Interaction A x B:
$\{(Xab - Xb) - (Xa - X)\} = \{(Xab - Xa) + (Xb - X)\}$

ANOVA: 2^2 FACTORIAL DESIGN

Model: DIET DRUG DIET*DRUG DIET*DRUG*subj(E)

SOURCE OF VARIATION	SUM OF SQUARES	df	MEAN SQUARE	F	p
Between groups					
TREATMENTS:					
DIET	3836.45	1	3836.45	16.08	0.001
DRUG	23188.05	1	23188.05	97.18	0.0001
DIET + DRUG	312.05	1	312.05	1.31	0.269
ERROR (GROUP X SUBJECT)	3817.6	16	238.6		
Total	31154.15	19			