



SEANCE  
REVISION  
BIOSTAT



# Tutorat Biostatistique 2012/2013

La pensée du jour :

« *La biostat' c'est bien, le tutorat de biostat' c'est encore mieux !!!* » A méditer...

La 1<sup>ère</sup> séance de biostat' vous a plu...

Alors vous allez sûrement encore plus adorer la seconde séance...



(enfin on l'espère...)

## LE SAVIEZ VOUS ???

Le Biostat' Style a fait son chemin depuis la dernière fois...

On nous a même organisé un voyage en Corée du Sud pour revenir sur les traces du Biostat Style...



# Quelques infos avant de commencer

- > Pas besoin de prendre de note.
- > Objectif : revoir les points importants.
- > Diaporama dispo après la séance.
- > **VIVE LA BIOSTAT !!!!**



# Séance de Révision Biostatistique

Episode 2 :

*The Eye Of  
The Biostat !*

# Déroulement de la séance

**I. Partie Probabilités (Pr. Staccini)**

**II. Partie Biostatistiques (Pr. Benoliel)**

**III. Partie Applications (Pr. Staccini/Lupi Pégurier)**

# Déroulement de la séance

Dans chaque partie :

**COMPREHENSION & REVISION !**

- Ce qu'il est important de se rappeler
- QCMs illustratifs des anciens concours
- Pièges facilement « piègeables »

# I. Partie Probabilités (Pr. Staccini)



**VIVE LA  
BIOSTAT !!!**

# Grand Récap sur les lois de Probas !

Loi de Bernoulli :  $P(X = k) = p^k q^{1-k}$

Loi Binomiale :  $P(X = k) = C_n^k p^k q^{n-k}$

Loi de Poisson :  $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$

Loi Géométrique :  $P(X = k) = pq^{k-1}$

Loi Hypergéométrique :  $P(X = k) = \frac{C_D^k * C_{N-D}^{n-k}}{C_N^n}$

Loi Exponentielle :  $P(X \leq x) = 1 - e^{-\lambda x}$

Loi Uniforme :  $P(X \leq x) = (x - a) / (b - a)$

Loi Normale centrée réduite :  $Z = (X - \mu) / \sigma$

# Grand Récap sur les approximations !

## 1/ De la loi Binomiale par la loi de Poisson :

si  $n > 50$        $p \leq 0,10$       ET       $np < 5$

$$\mathcal{B}(n; p) \rightarrow \mathcal{P}(\lambda = np)$$

## 2/ De la loi Binomiale par la loi Normale :

si       $np \geq 5$       et       $nq \geq 5$

$$\mathcal{B}(n;p) \rightarrow \mathcal{N}(np; \sqrt{npq})$$

## 3/ De la loi de Poisson par la loi Normale :

si  $\lambda > 25$        $\mathcal{P}(\lambda) \rightarrow \mathcal{N}(\lambda; \sqrt{\lambda})$

# QCM Loi normale centrée réduite

Concernant l'IMC, on a les différentes classes suivantes :

Classification	Maigreur extrême	Maigreur	Poids normal	Surpoids	Obésité	Obésité morbide
IMC	<16	[16-18,5[	[18,5-25[	[25-30[	[30-40[	>40

On a mené une étude concernant la répartition de la population dans ces différentes classes, et on remarque que l'IMC de la population peut être assimilé à une loi normale  $N(25;8)$ . Quelle est la proportion de sujets classés dans la catégorie « poids normal » ?

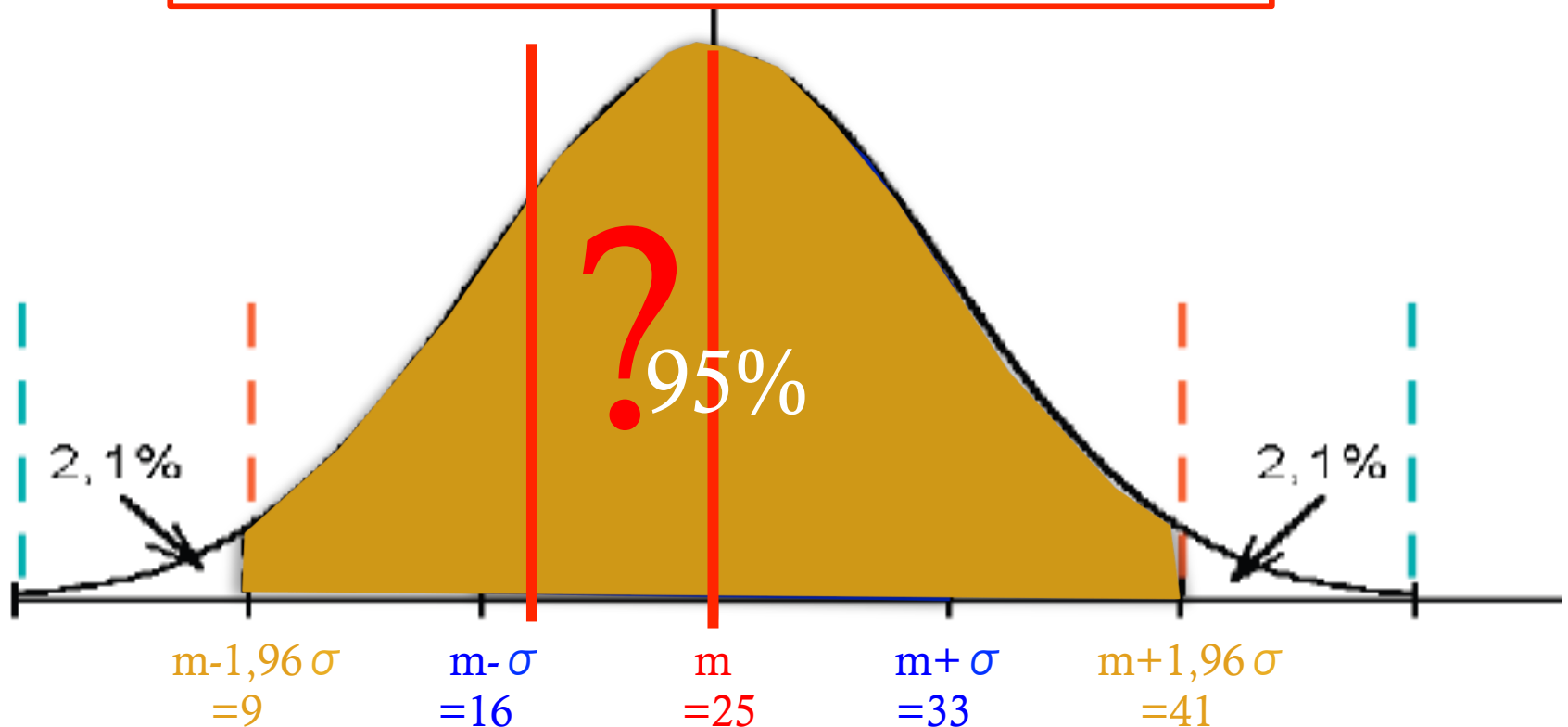
u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7290	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9779	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

# Pour aller plus loin...

## L'estimation

$[\mu - \sigma ; \mu + \sigma]$  contient 68% de la population

$[\mu - 1,96 \sigma ; \mu + 1,96 \sigma]$  contient 95% de la population



# QCM Loi normale centrée réduite

Comment fait-on pour trouver  $P(x \text{ appartient } [18,5 - 25])$  ?

C'est tout bêtement  $P(x < 25) - P(x < 18,5)$

On passe par la loi normale centrée réduite :

$$P(Z < (25 - 25) / 8) - P(Z < (18,5 - 25) / 8)$$

$$P(Z < 0) = 0,5 \text{ dans la table}$$

$$P(Z < -6,5 / 8) = P(Z < -0,81) = 1 - P(Z < 0,81)$$

$$= 1 - 0,79 = 0,21$$

$$0,5 - 0,21 = 0,29 = 29\%$$

# QCM Concours UE4

**QCM 1 : Un contrôle mensuel du bon état des instruments chirurgicaux est effectué dans chaque bloc opératoire. Lors du dernier contrôle portant sur un lot de 100 instruments, 4 sont retrouvés défectueux. Durant le mois précédant le contrôle, 100 interventions différentes indépendantes ont été réalisées. A chaque intervention, le chirurgien n'a utilisé qu'un seul des 100 instruments du lot contrôlé. La probabilité pour que le chirurgien n'ait utilisé aucun instrument défectueux durant cette période.**

- A) est égale à  $0,04^{100}$ .
- B) est égale à  $1 - 0,96^{100}$ .
- C) est égale à  $0,96^{100}$ .
- D) est donnée par la loi binomiale de paramètres  $n=100$  et  $p=0,04$ .
- E) Les propositions A, B, C et D sont fausses.

**CORRECTION : C, D**

# QCM Concours UE4

Il y a deux résultats possibles pour chaque intervention : Succès : instrument qui fonctionne - Echec : instrument qui ne fonctionne pas.

Application de la loi binomiale:  $B(n = 100 ; p = 0,04)$

$$\begin{aligned} P(X = k = 0 \text{ instruments défectueux}) &= C_n^k p^k (1-p)^{(n-k)} \\ &= C_{100}^0 0,04^0 (1 - 0,04)^{(100-0)} = C_{100}^0 0,04^0 0,96^{100} = 1 \times 0,96^{100} \\ &= 0,96^{100} \end{aligned}$$

# QCM Concours UE4

**QCM 3 : Le nombre de rhumes attrapés par un individu en l'espace d'un an est une variable aléatoire de Poisson de paramètre  $\lambda = 4$ . Un remède miracle (basé sur l'effet de la vitamine C à haute dose) a été lancé sur le marché. Il abaisse le paramètre  $\lambda$  à 1 pour 75% de la population. Pour le reste de la population, le remède n'a pas d'effet appréciable. Si 200 individus essaient le remède miracle, le nombre moyen de rhumes qu'ils attrapent durant une année.**

A) est égal à 150

B) est égal à 200

C) est égal à 350

D) est donné en appliquant une loi de Poisson de paramètre  $\lambda = 1+4 = 5$

E) est donné en appliquant une loi de Poisson de paramètre  $\lambda = 1/4 = 0,25$

**CORRECTION : C**

# QCM Concours UE4

Différent du cours → Réfléchir à ce qu'il faut faire → en l'occurrence simplement de la logique, pas de formule ! 😊

On sait que dans la Loi de Poisson,  $\lambda = \mu = \sigma^2$  (c'est ce qui traduit le caractère poissonien de la variable, youpi...). Tout ça pour dire que le lambda, c'est la moyenne de rhumes !

Le nombre moyen de rhumes pour toute la population est :

$$(75\% \times 200) \times 1 \text{ rhume} + (25\% \times 200) \times 4 \text{ rhumes}$$

$$= 150 \times 1 + 50 \times 4 = 350 \text{ rhumes}$$

## II. Partie Biostatistique (Pr. Benoliel)



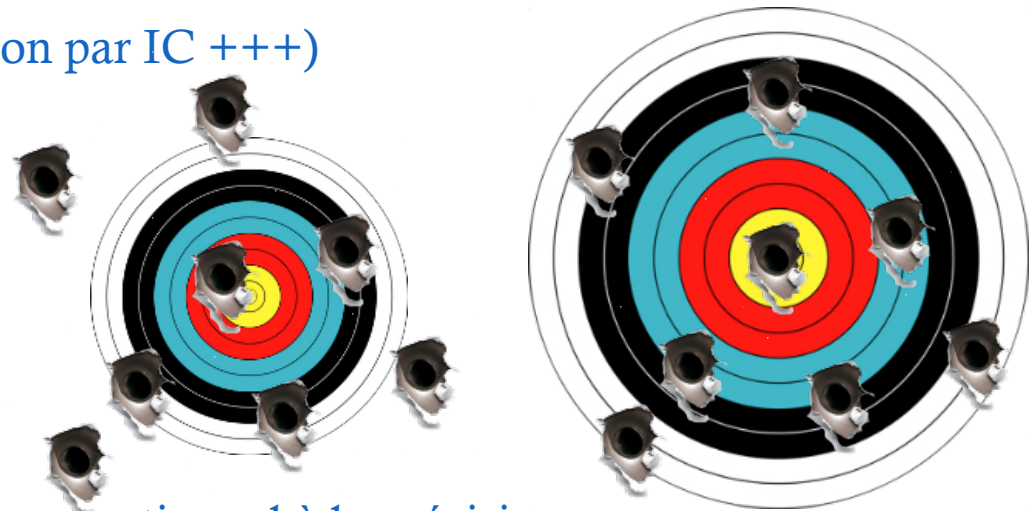
**VIVE LA  
BIOSTAT !!!**

# Grand Récap sur les généralités !

Notion de **Variabilité**, fondamentale (inter et intra individuelle)  
**Paramètres** (position et dispersion)

**Estimation par intervalle** (comparaison par IC +++)  
**Représentativité assurée par TAS**

+  $\alpha$  est petit,  
+  $\varepsilon$  est grand,  
→ plus l'IC est grand.  
→ Plus la précision est petite



!!! Indice de précision (i), inversement proportionnel à la précision

	Rejet H0	Non-Rejet H0
H0 vraie		
H0 fausse = H1 vraie		

# Grand Récap sur l'estimation !

## Données quantitatives

Moyenne :  $m$   
Ecart-Type :  $s$

Estimation de la moyenne  
inconnue dans la population :

$$\mu \in \left[ m \pm \frac{\varepsilon s}{\sqrt{n}} \right]$$

## Données qualitatives

% au niveau de la population :  $p_0$   
Ecart-Type :  $s$

Estimation du pourcentage  
inconnu dans la population :

$$p \in \left[ p_{obs} \pm \varepsilon s \right]$$

$\varepsilon$  lu dans la table  $\rightarrow$  risque d'erreur accepté

# Grand Récap sur les tests !

But : tester hypothèses

H0 : pas de différence

H1 : différence significative

Etapas :

1) Définir hypothèses

2) Déterminer variables

3) Choix du test adapté

4) Choix risque d'erreur

5) Recueil données et Z

6) Interprétation

Effectif	Quantitatives	Qualitatives	Qualitatives-Quantitatives
>4 & <12			
>12 & <30			
>30			

Valeur calculée > Valeur théorique → repousse H0 (=accepte H1)

Valeur calculée < Valeur théorique → accepte H0 (=repousse H1)

(piège : U Mann & Withney le seul test faisant l'inverse...)

# QCM Concours UE4

**QCM 1 : Un rhumatologue pense que l'apparition de rhumatismes est associée au poids des patients. Pour tenter de le démontrer, il décide de constituer deux groupes de patients : un groupe de patients souffrant de rhumatismes (groupe R) et un groupe de patients ne souffrant d'aucun rhumatisme (groupe SR). Ensuite, il compare la moyenne du poids des patients dans le groupe R à la moyenne du poids des patients dans le groupe SR.**

- A) Cette étude s'apparente à une enquête de cohorte
- B) La formulation de l'hypothèse  $H_0$  nécessite de fixer un risque de première espèce le plus faible possible
- C) L'hypothèse  $H_1$  est : la moyenne du poids des patients du groupe R est comparable à la moyenne du poids des patients du groupe SR
- D) L'hypothèse  $H_0$  est : la moyenne du poids des patients dans le groupe R est différente de la moyenne du poids des patients dans le groupe SR
- E) **CORRECTION : E** Le nombre de sujets à inclure dans l'étude dépend en partie de la variance de la mesure du poids dans chaque groupe

# QCM Concours UE4

- A) Faux. On revient sur leur passé, on ne va pas vers l'avant c'est plus une étude de type cas-témoins.
- B) Faux. Le risque de première espèce est fixé à priori sans rapport avec les hypothèses.
- C) Faux. C'est la définition de  $H_0$ .
- D) Faux. C'est la définition de  $H_1$ .
- E) Vrai. C'est un des paramètres qui entre en jeu tout comme la différence à mettre en évidence...

# QCM Concours UE4

**QCM 2-3 : On souhaite évaluer l'intérêt d'une substance S capable de désintoxiquer les fumeurs. On randomise deux groupes de 50 fumeurs choisis au hasard. Le premier groupe reçoit la substance S, l'autre reçoit un placebo P. Le traitement dure deux mois pour les deux groupes. La consommation de cigarettes (mesurée en nombre de cigarettes par jour) est relevée avant et après traitement pour chaque patient dans chaque groupe. Le tableau ci-dessous donne la moyenne de la consommation ( $\bar{x}_n$ ) et la variance ( $s^2$ ) pour chaque groupe.**

# QCM Concours UE4

	S (n=50)		P (n=50)	
	m	s <sup>2</sup>	m	s <sup>2</sup>
Consommation avant traitement	19,5	54,2	16,5	35,6
Consommation après traitement	5,4	30,4	3,8	20,1
Variation de consommation	14,1	9,1	12,7	8,9

# QCM Concours UE4

**QCM 2 : On cherche à savoir si les consommations moyennes de cigarettes avant traitement sont comparables entre les deux groupes.**

- A) Il s'agit d'étudier l'existence d'une relation entre deux variables qualitatives.
- B) Il s'agit d'étudier l'existence d'une relation entre une variable qualitative et une variable quantitative.
- C) L'hypothèse  $H_0$  est : il n'y a pas de différence de consommation avant traitement entre les deux groupes.
- D) Pour répondre à la question, on peut utiliser un test du Chi-2.
- E) Pour répondre à la question, on ne peut pas utiliser un test de comparaison de moyennes.

**CORRECTION : B, C**

# QCM Concours UE4

- A) Faux. On compare la consommation de cigarette (quantitative), dans les 2 groupes (qualitative) avant le traitement.
- B) Vrai. Même justification.
- C) Vrai.
- D) Faux. Test du Chi-2 pour deux variables qualitatives !
- E) Faux. Oui on peut l'utiliser car effectif  $>30$  et variable qualitative et quantitative...

# QCM Concours UE4

**QCM 3 : On cherche à savoir si les deux groupes diffèrent pour la variation de la consommation de cigarettes après traitement.**

- A) Il s'agit d'étudier l'existence d'une relation entre deux variables quantitatives.
- B) Il s'agit d'étudier l'existence d'une relation entre une variable quantitative et une variable qualitative.
- C) L'hypothèse  $H_0$  est : il y a une différence de consommation entre les deux groupes après traitement.
- D) On réalise le test statistique adapté. Le  $\varepsilon$  calculé = 2,3. On conclut à l'efficacité de la substance S, au risque de 5%.
- E) On réalise le test statistique adapté. Le  $\varepsilon$  calculé = 2,3. On conclut à l'absence d'efficacité de la substance S.

**CORRECTION : B, D**

# QCM Concours UE4

- A) Faux. On compare la variation de consommation de cigarette (quantitative), dans les 2 groupes (qualitative) après le traitement.
- B) Vrai. Même justification.
- C) Faux. C'est la définition de H1.
- D) Vrai. Le test adapté est celui de comparaison de moyenne.  
Or on sait que  $\varepsilon$  calculé = 1,6 et  $\varepsilon$  théorique = 2,3.  
Donc  $\varepsilon$  calculé >  $\varepsilon$  théorique  $\Rightarrow$  On repousse H0 = accepte H1.  
Etude menée dans les règles de l'art et variation en faveur de la substance.
- E) Faux.

# Les pièges !!!

- Dans la symétrie des hypothèses
- Bien manipuler  $H_0/H_1$  /accepter/repousser
- U Mann & Withney, seul test ne respectant pas la règle (l'inverse)
- Vérifier le représentativité de l'échantillon
- Possibilité d'extrapoler en disant plus d'efficacité si échantillon représentatif ET résultats en faveur de cela

# III. Partie Applications (Pr. Staccini/Lupi Pégurier)



**VIVE LA  
BIOSTAT !!!**

# Grand Récap sur les tests !

	MALADES	NON MALADES
T+	VP	FP
T-	FN	VN

# Grand Récap sur Se/Sp/VPP/VPN !

La sensibilité et la spécificité sont PROPRES AU TEST et donc INDEPENDANTES de la prévalence de la maladie dans la population → Paramètres INTRINSEQUES!

VPP et VPN dépendent de la prévalence de la maladie → Paramètres EXTRINSEQUES du test!

# Grand Récap sur la Sensibilité !

	<b>MALADES</b>	<b>NON MALADES</b>
T+	VP	FP
T-	FN	VN

**SENSIBILITE = LECTURE VERTICALE DU  
TABLEAU 1ERE COLONNE =  $VP / (VP+FN)$**

# Grand Récap sur la Spécificité !

	MALADES	NON MALADES
T+	VP	FP
T-	FN	VN

**SPECIFICITE = LECTURE VERTICALE DU  
TABLEAU 2EME COLONNE =  $VN/(VN+FP)$**

# Grand Récap sur la VPP !

	MALADES	NON MALADES
T+	VP	FP
T-	FN	VN

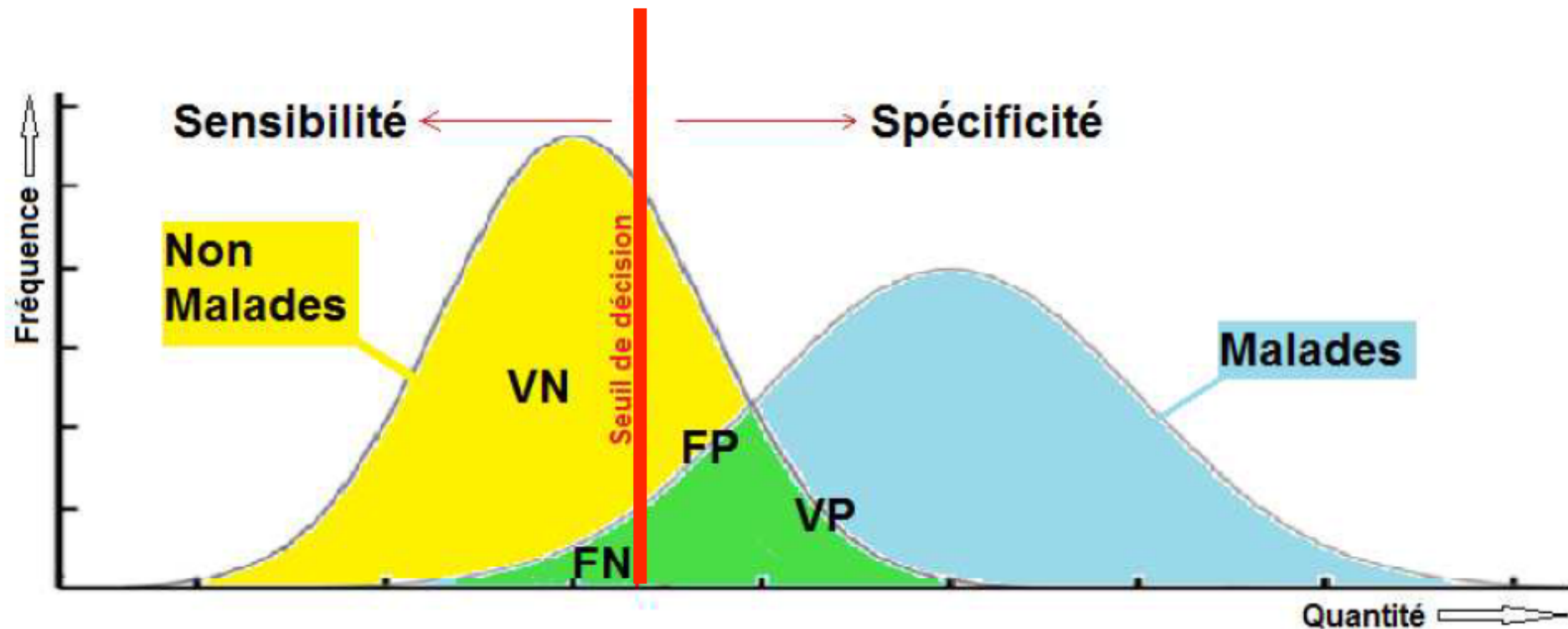
**VPP = LECTURE HORIZONTALE DE LA 1ERE  
LIGNE DU TABLEAU:  $VPP = VP / (VP + FP)$**

# Grand Récap sur la VPN !

	MALADES	NON MALADES
T+	VP	FP
T-	FN	VN

**VPN = LECTURE HORIZONTALE DE LA 2EME  
LIGNE:  $VPN = VN / (FN + VN)$**

# Grand Récap sur la Courbe ROC !

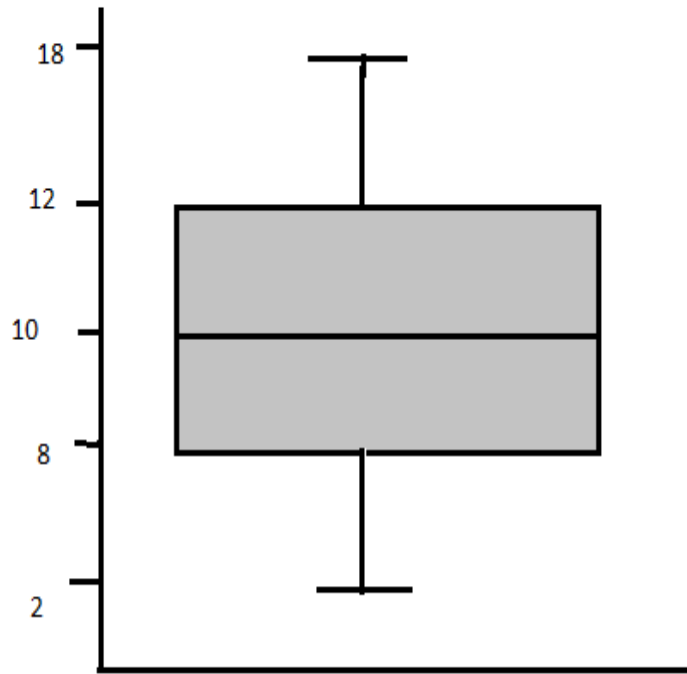


SEUIL ↓ SENSIBILITE ↑ → VP ET FP ↑ → SPECIFICITE ↓

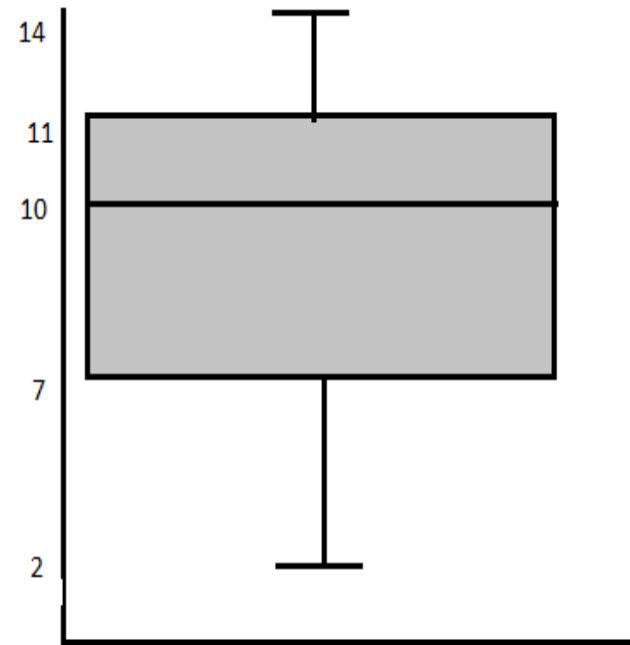
SEUIL ↑ SPECIFICITE ↑ → VN ET FN ↑ → SENSIBILITE ↓

# Grand Récap sur la Box Plot !

SYMETRIQUE



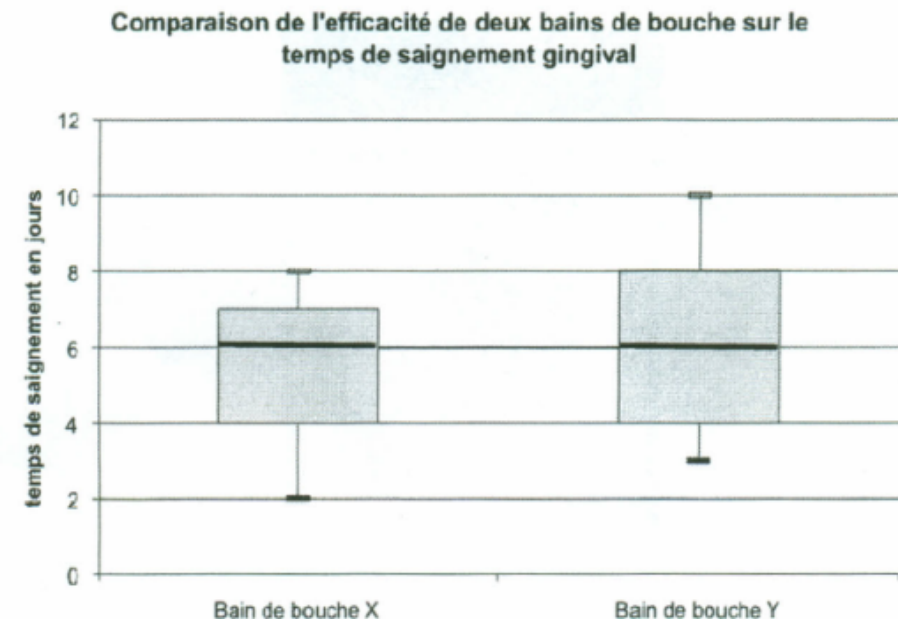
ASYMETRIQUE



# QCM Concours UE4

**QCM 1 : Une étude clinique a évalué l'efficacité de deux bains de bouche sur les saignements gingivaux. La comparaison des deux produits est représentée par un boxplot (boîte à moustaches) ci-dessous.**

- A) Les durées moyennes de saignement sont équivalentes avec les deux bains de bouche.
  - B) La distribution des durées de saignement dans chaque groupe est symétrique.
  - C) 50% des patients ont saigné entre 4 et 8 jours avec le bain de bouche Y.
  - D) 25% des patients ont saigné moins de 4 jours avec les deux bains de bouche.
  - E) Les propositions A, B, C, D sont fausses.
- CORRECTION : C, D**



# QCM Concours UE4

- A) Faux. Sur les box-plots--> MOYENNE NON INDIQUEE
- B) Faux. Visible au premier coup d'œil.
- C) Vrai. Distance interquartile :  $Q3-Q1 = 50\%$  des gens.
- D) Vrai. Personnes en dessous de  $Q1 = 25\%$ .
- E) Faux.

# QCM Concours UE4

**QCM 2 : Un examen diagnostique révèle que 1% des sujets de plus de 55 ans présentent des bactéries pathogènes pour le parodonte (tissus de soutien de la dent : os, gencive, ligament). Parmi ces sujets, 60% se révèlent être effectivement atteints d'une parodontopathie. Par ailleurs, on considère que la prévalence des sujets porteurs de parodontopathies est de 4% dans cette population. Sur une population de 10 000 sujets, on peut dire que :**

- A) La sensibilité du critère "présence de bactéries pathogènes" est 15%.
- B) La sensibilité du critère "présence de bactéries pathogènes" est 20%.
- C) La sensibilité du critère "présence de bactéries pathogènes" est 85%.
- D) Le nombre de vrais négatifs est de 9560.
- E) La valeur prédictive positive du test est 60%.

**CORRECTION : A, D**

# QCM Concours UE4

	MALADES	NON MALADES	TOTAL
B+	60	40	100
B-	340	9560	9900
TOTAL	400	9600	10000

A) Vrai. Sensibilité= $60/400=3/20=0,15=15\%$

B) Faux

C) Faux

D) Vrai

E) Faux

C'est OK ?




Le Tutorat de Biostat' c'est déjà fini ☹ ☹ ☹...

Mais RDV sur le forum pour la dernière ligne droite et le jour du concours pour vous encourager !!!



Vive la  
biostat' !



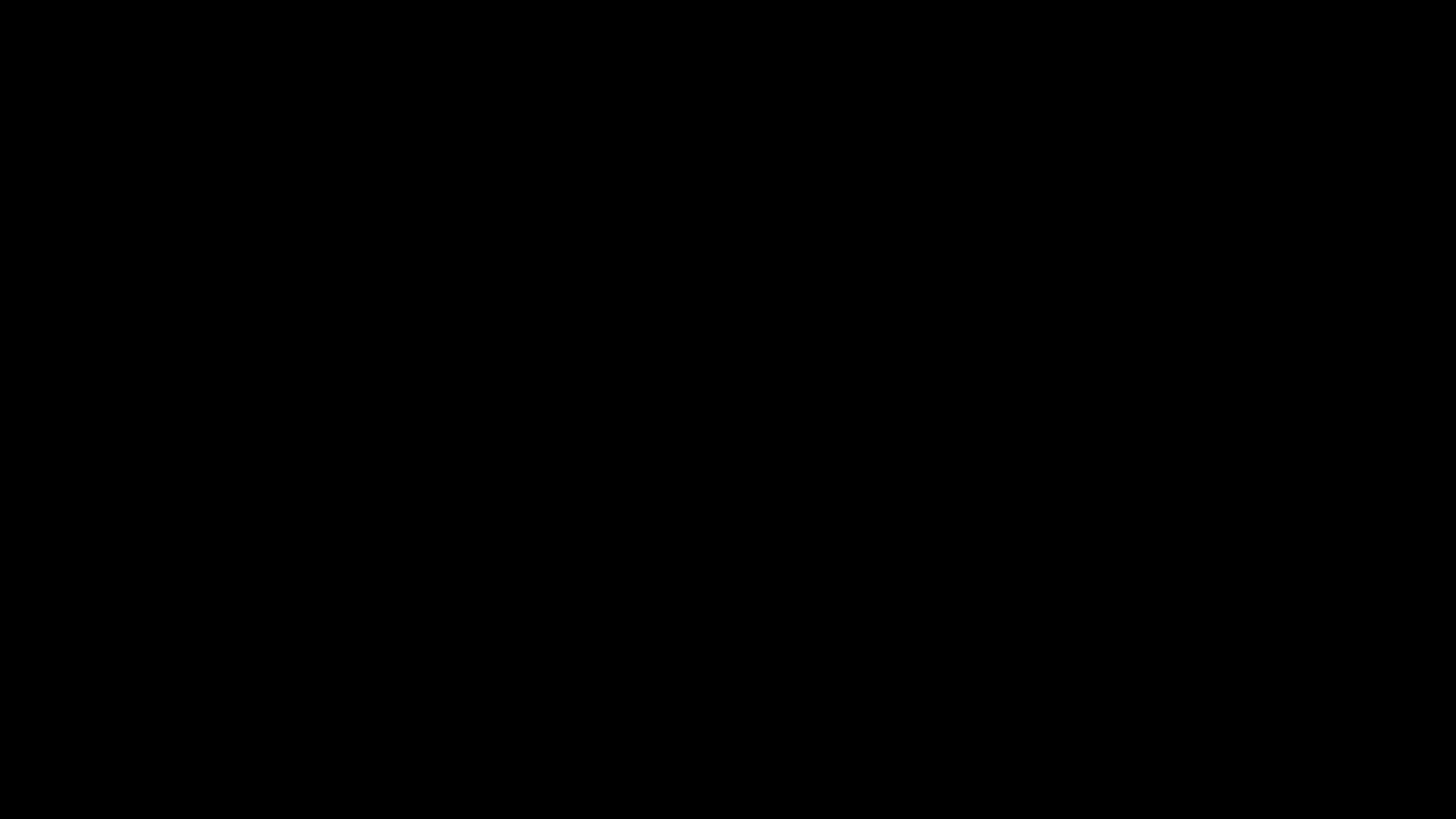
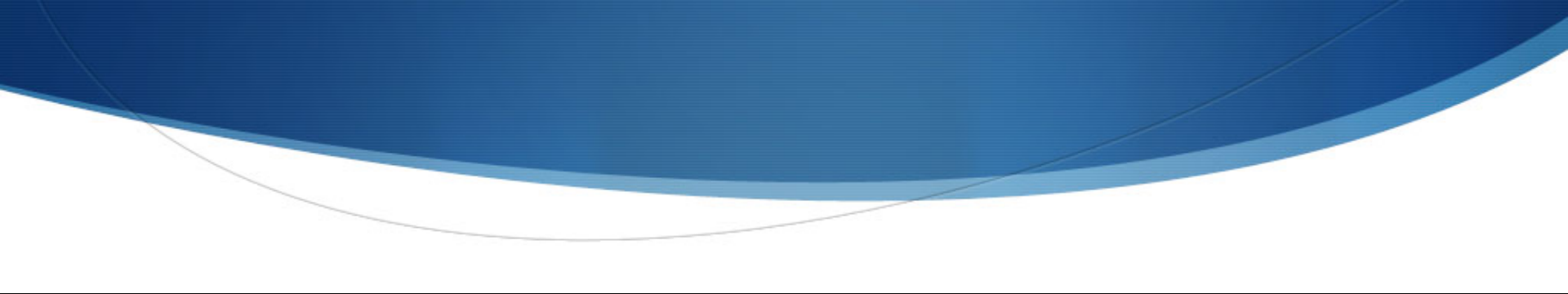
*et.....*



Finissons la séance

par une chanson,

Pour être de bonne humeur...



FAUX !!!



# Biostat Rock Anthem !



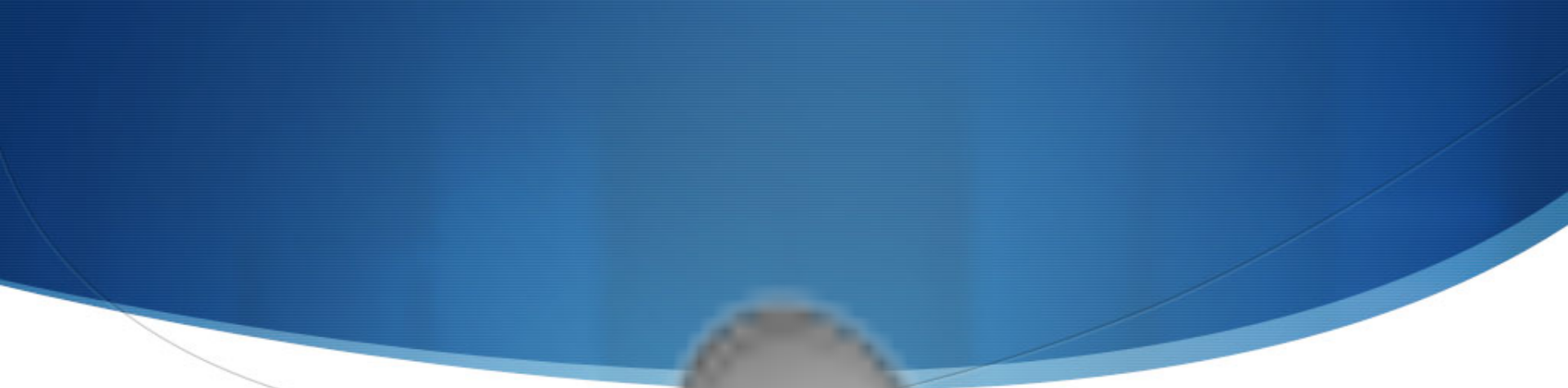
Merci à tous pour votre attention !

Travaillez bien à fond,  
posez toutes vos questions sur le forum !!!

A l'année prochaine en P2 !!!

---

*Tous droits réservés au tutorat niçois. Le tutorat est gratuit.  
Toute reproduction ou vente est interdite.*



On est tous derrière vous !!!



**TURORAK 2012-2013**

# LE BIOSTAT' STYLE !



