

Tutorat Biostatistique 2012/2013

La pensée du jour :

« *La biostat' c'est une philosophie, si tu l'aimes tout va bien !!!* » A méditer...

Les cours de biostat' vous ont plu...

Alors vous allez sûrement adorer la
séance de révision biostatistique



(enfin on l'espère...)

Avant de commencer....

Merci d'applaudir le
professeur Benoliel pour
sa participation active au
tutorat (relecture sujet,
réponses, présence
séance...)



Quelques infos avant de commencer

- > Pas besoin de prendre de note.
- > Objectif : dissiper vos moindres doutes.
- > Diaporama dispo après la séance (pas les QCM).
- > **VIVE LA BIOSTAT !!!!**



Séance de Révision Biostatistique

Episode 1 :

La biostat'

ne meurt jamais !

Déroulement de la séance

I. Réponses aux questions

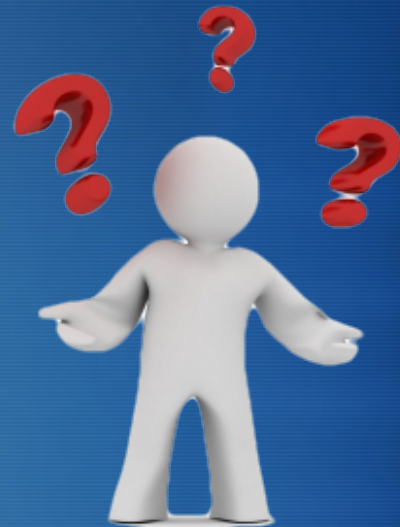
II. Reprise de notions

III. QCM Time

III. Correction des QCMs

IV. Le grand récap'

REPONSES AUX QUESTIONS !!!



Vos questions, ses réponses...



I-1 Réponses aux questions

QUESTION :

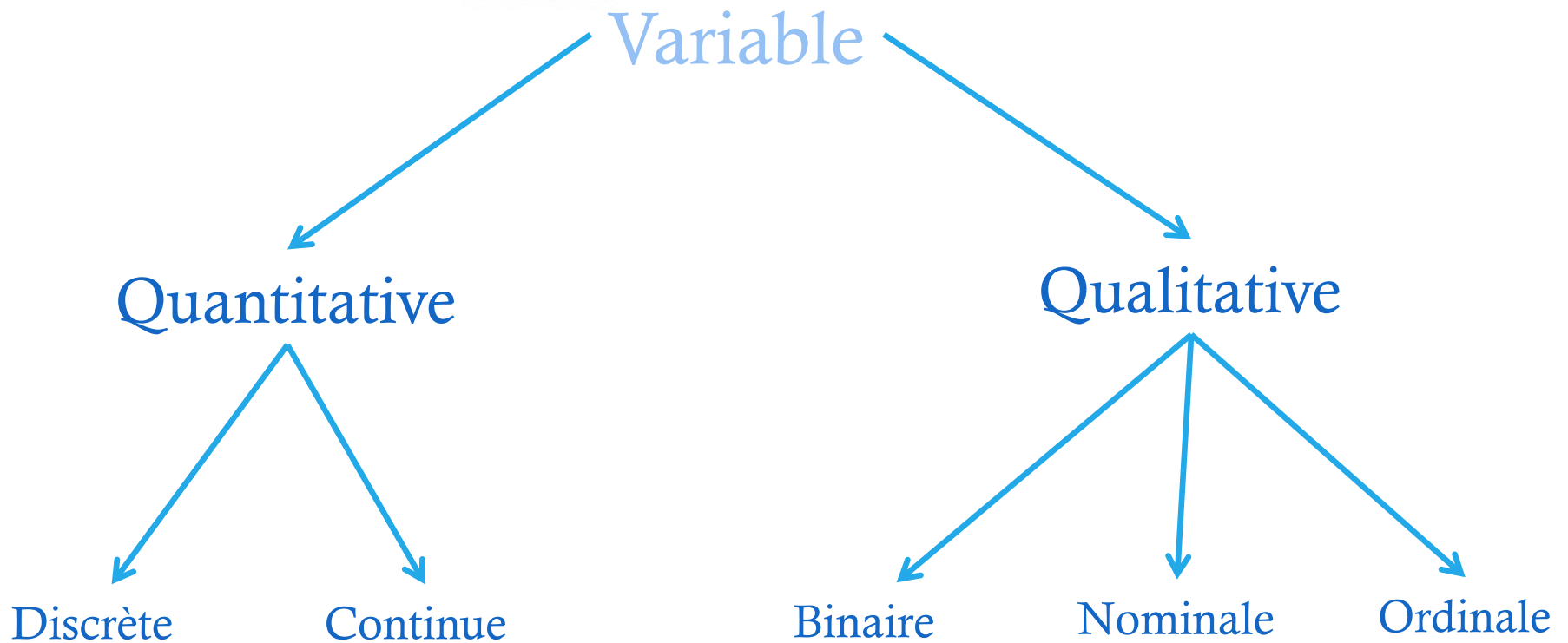
« La variable qualitative **binaire**, est-elle une **sous-catégorie** de la variable qualitative nominale ?

Ou une **3^{ème} catégorie** de variable qualitative ?

Avis unanime avec le **professeur Staccini** ?

Que répondre au QCM ? »

II-1 Les variables



II-1 Les variables

QUESTION :

« La variable qualitative **binaire**, est-elle une **sous-catégorie** de la variable qualitative nominale ?

Ou une **3^{ème} catégorie** de variable qualitative ?

Avis unanime avec le **professeur Staccini** ?

Que répondre au QCM ? »

REPONSE : Le type le plus simple de variable qualitative est le type « binaire », à **2 modalités**. Si il y a plus de 2 modalités, on parle de type « nominal ».

En ce sens on peut dire que « **binaire** » est une **sous catégorie** de « nominale ».

I-2 Réponses aux questions

QUESTION :

« A quoi correspond l'abscisse de la courbe de Gauss ? »

I-2 Réponses aux questions

La **loi de Gauss** est une loi très utilisée en statistique car elle permet de décrire une population, dont l'effectif est supérieur à 30 ($n > 30$).

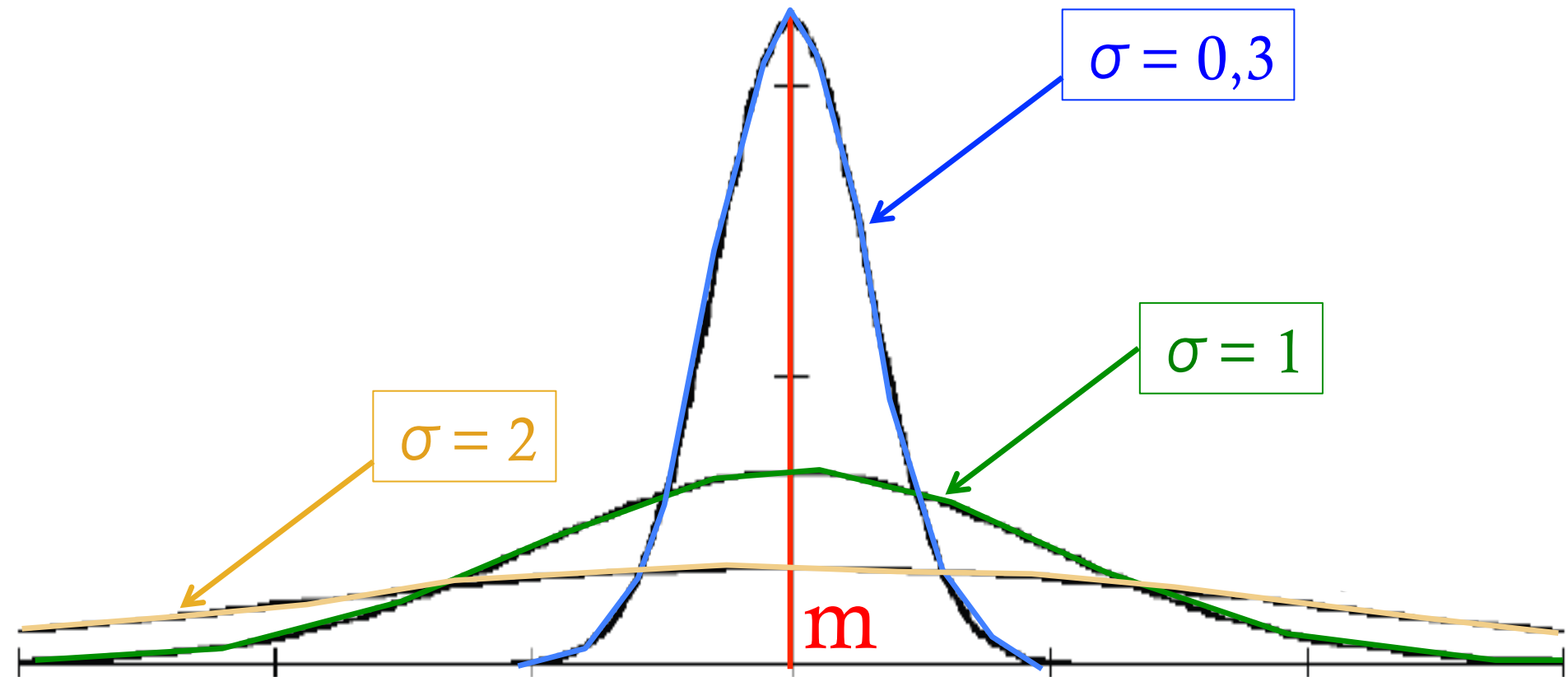
Elle permet de visualiser :

-moyenne

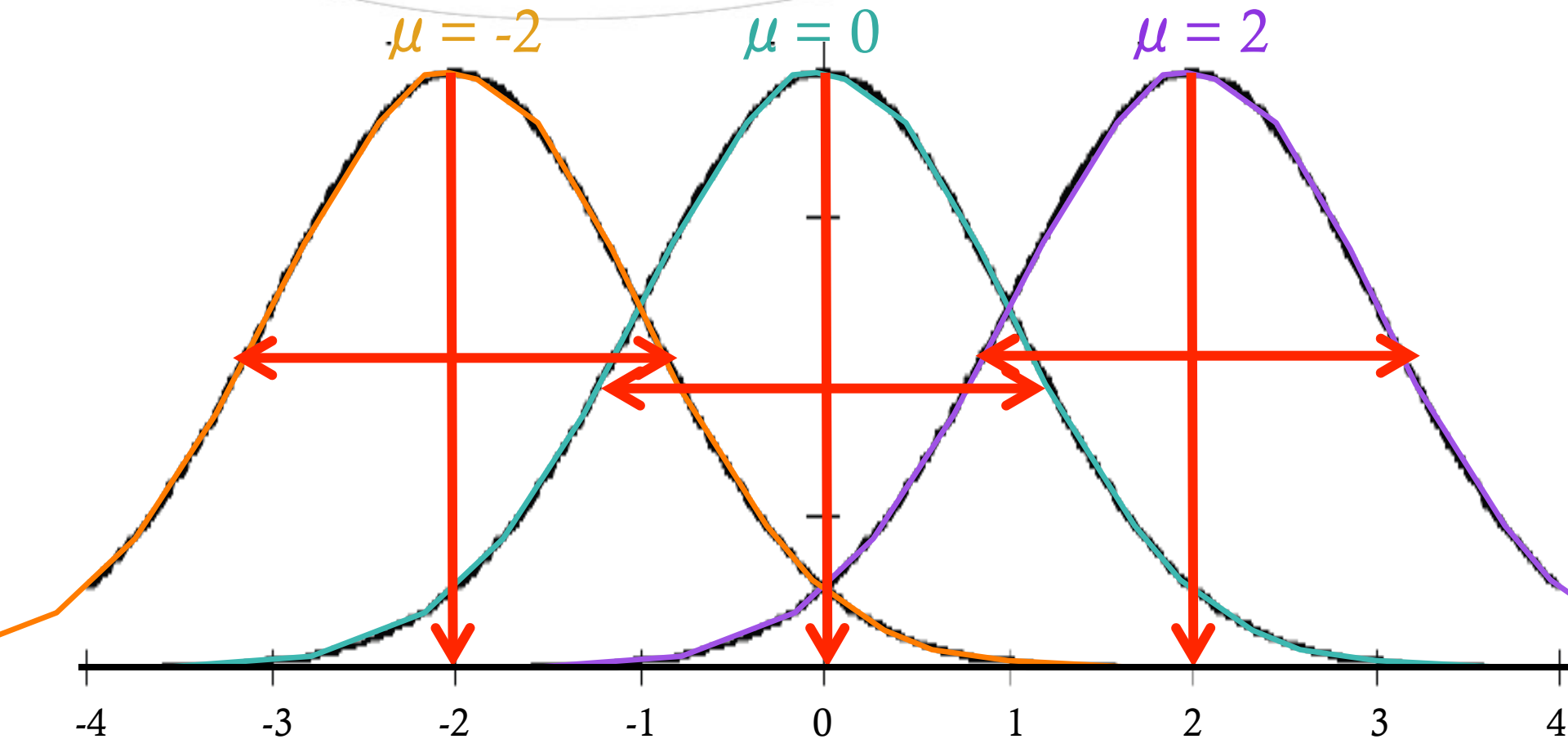
-écart-type

-intervalle de confiance

I-2 Réponses aux questions

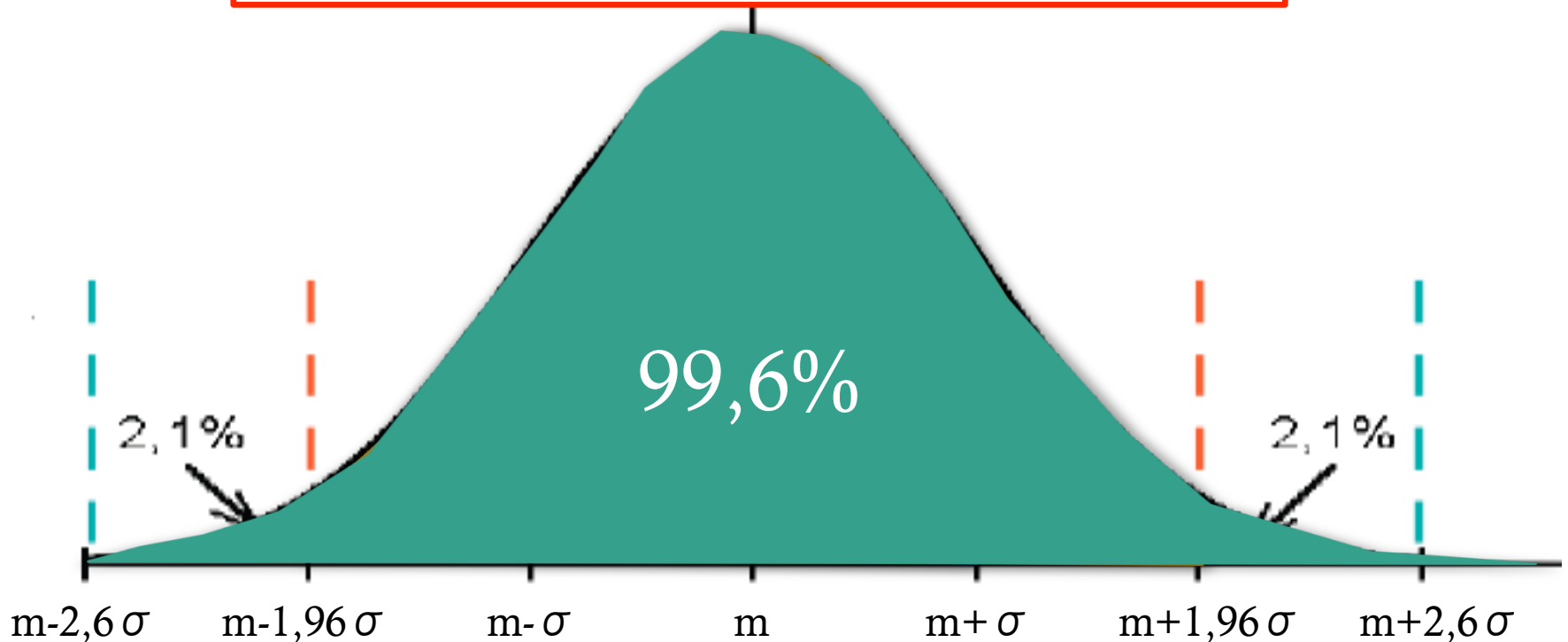


I-2 Réponses aux questions



I-2 Réponses aux questions

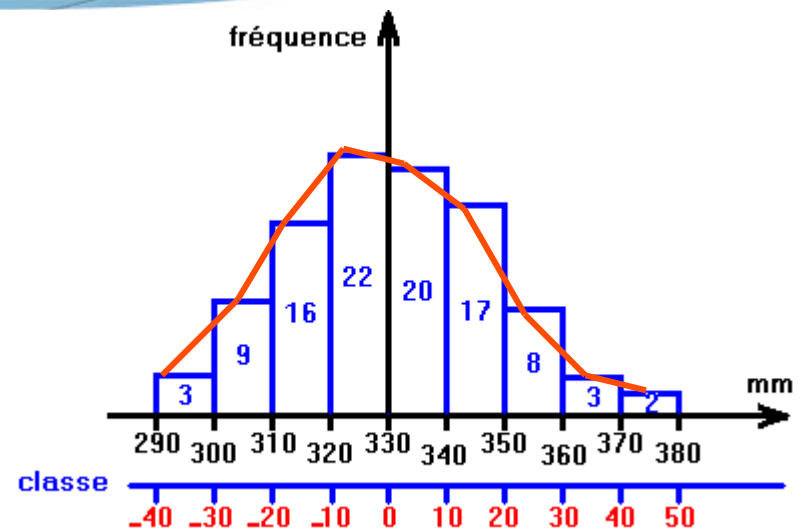
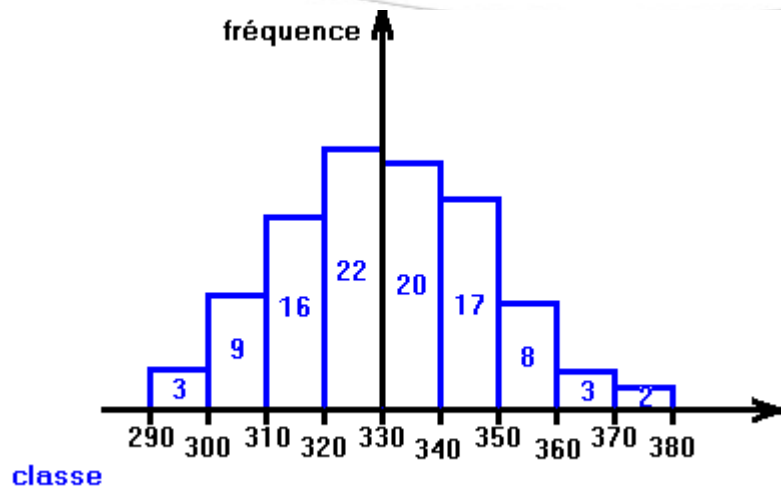
$[\mu - \sigma ; \mu + \sigma]$ contient 68,2% de la population
 $[\mu - 1,96 \sigma ; \mu + 1,96 \sigma]$ contient 95,4% de la population
 $[\mu - 2,6 \sigma ; \mu + 2,6 \sigma]$ contient 99,6% de la population



I-2 Réponses aux questions



I-2 Réponses aux questions



Ex : Histogramme identique, avec abscisses relatives à la moyenne

En généralisant, si on fait **tendre** le nombre de mesures vers l'**infini** et la **largeur** des classes vers **zéro**, cette courbe **tend** vers une courbe **limite** appelée "**courbe de**

Gauss", d'équation

$$y = \frac{1}{\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

I-2 Réponses aux questions

QUESTION :

« A quoi correspond l'**abscisse** de la courbe de **Gauss** ? »

REPONSE :

La courbe de Gauss représente la loi de probabilité « normale centrée réduite ». Elle représente de manière théorique une **distribution probabiliste bien particulière**. En fait, cette distribution est utilisée pour définir **une marge d'erreur** lors de l'estimation de la moyenne. On parle de

l'erreur standard de la moyenne $\sigma = s/$

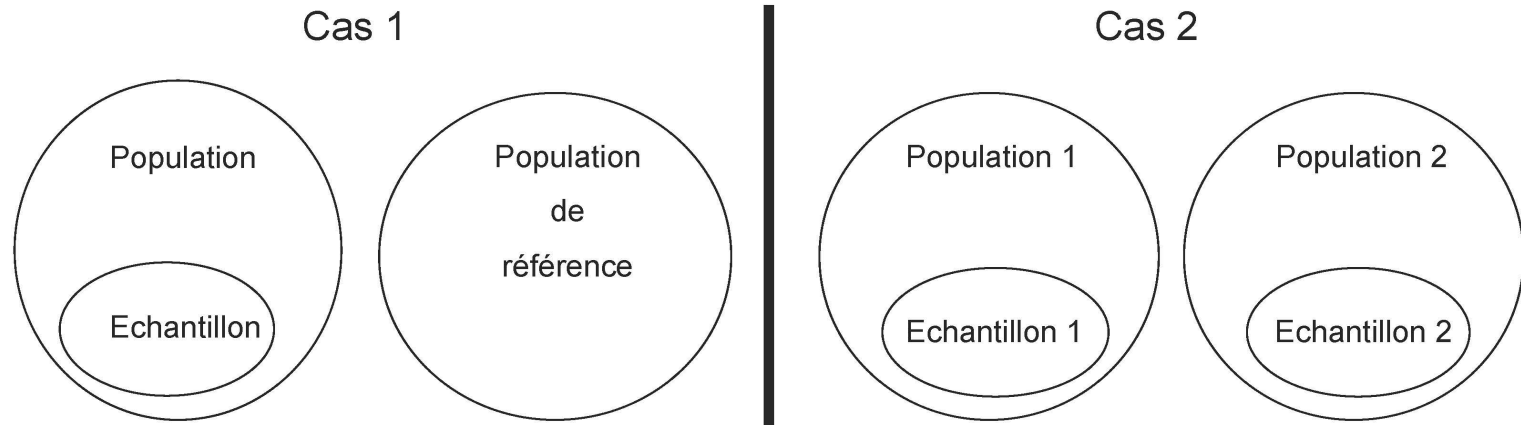
I-3 Réponses aux questions

QUESTION :

« Dia8-9 du cours 3 : Pouvez vous nous expliquer pourquoi peut on comparer une population à une **population** dites de **référence** sans **échantillonner** cette dernière ?»

I-3 Réponses aux questions

Les tests de comparaison : 2 cas



Différence observée :

Fluctuations d'échantillonnage ?

Différences réelles entre populations ?

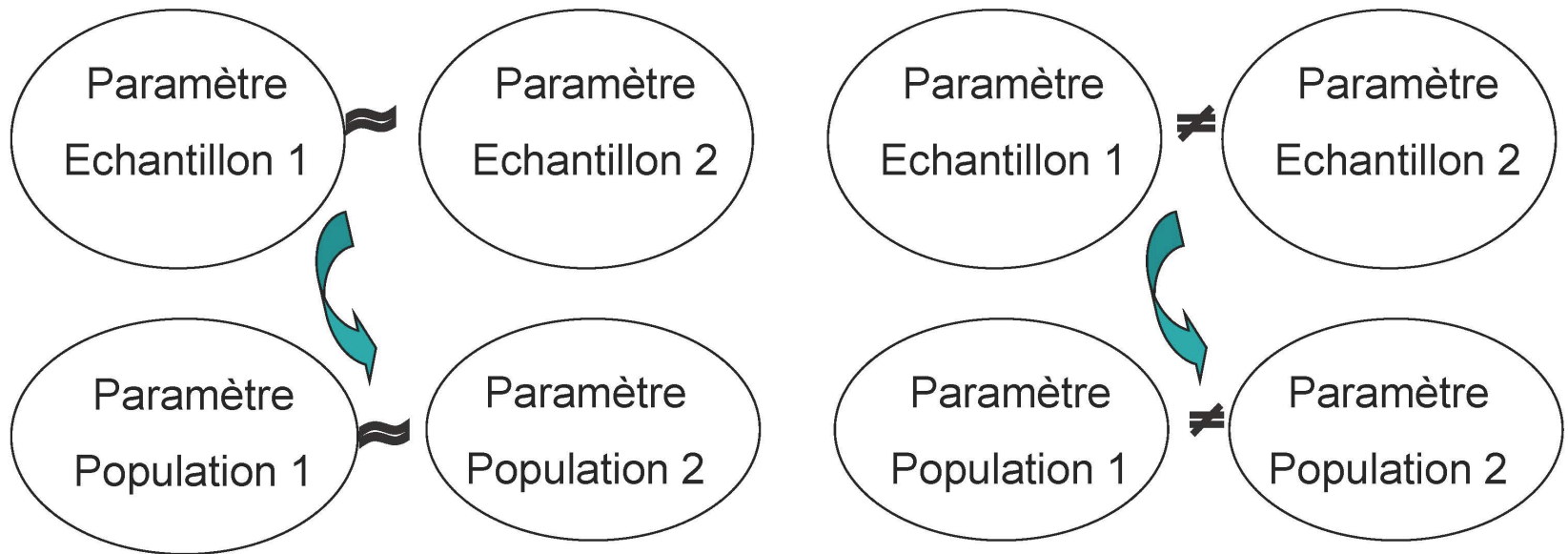
Hasard?



Test Statistique

I-3 Réponses aux questions

Formulation des hypothèses



Hypothèse nulle (H0)

Hypothèse alternative (H1)

I-3 Réponses aux questions

QUESTION :

« Dia8-9 du cours 3 : Pouvez vous nous expliquer pourquoi peut on comparer une population à une **population** dites de **référence sans échantillonner** cette dernière ?»

REPONSE :

Par définition, une population de référence est une population qui est **parfaitement définie**, par les critères pris en compte dans l'étude ainsi que par son effectif, connu et fini. Dans ce cas, il **vaut mieux utiliser la population dans son entier** qu'un échantillon.

I-4 Réponses aux questions

QUESTION :

« Pouvez vous nous repréciser la distinction entre la **précision** et **l'indice de précision** ?

En **QCM**, les notions seront bien noté comme cela ? »

I-4 Réponses aux questions

$$\mu \in \left[m \pm \frac{\varepsilon S}{\sqrt{n}} \right] \Rightarrow \text{Intervalle au risque } \alpha$$

☞ Attention aux notations !

Dans l'échantillon :

n : effectif

m : moyenne

s : écart-type

Dans la population :

N : effectif

μ : moyenne

σ : écart-type

$$\alpha = 5\% \rightarrow \varepsilon = 1,96$$

$$\alpha = 1\% \rightarrow \varepsilon = 2,6$$

I-4 Réponses aux questions

$$\mu \in \left[m \pm \frac{\varepsilon S}{\sqrt{n}} \right] \Rightarrow \text{Intervalle au risque } \alpha$$

☞ Calcul de la valeur de la précision (=largeur de l'IC)

i =

$$\begin{aligned} \alpha = 5\% &\rightarrow \varepsilon = 1,96 \\ \alpha = 1\% &\rightarrow \varepsilon = 2,6 \end{aligned}$$

I-4 Réponses aux questions

$$\mu \in \left[m \pm \frac{\varepsilon S}{\sqrt{n}} \right] \Rightarrow \text{Intervalle au risque } \alpha$$

+ n est grand,
+ i (indice de la précision) est petit,
→ Meilleure est la précision

Indice de la précision quantifie la précision !!!

I-4 Réponses aux questions

QUESTION :

« Pouvez vous nous repreciser la distinction entre la **precision** et l'**indice de precision** ?
En **QCM**, les notions seront bien notées comme cela ? »

REPONSE : La **precision** est une **notion generale** qui peut être définie comme « bonne », « mauvaise »,... Afin d'être précis on a besoin de **quantifier cette notion en calculant un indice** $i = \frac{\epsilon S}{\sqrt{n}}$
Ex : la taille d'un individu (petite, haute,..).

L'indice que l'on comprend mieux est la valeur (1,75 m).

Il ne PEUT pas y avoir d'ambigüité !

I-5 Réponses aux questions

QUESTION :

« Pourquoi dans la technique des **séries appariés**, on utilise la **formule** du test **comparaison de moyennes** ?

Pourquoi on utiliserait pas une **autre formule** spécifique ? »

I-5 Réponses aux questions

QUESTION :

« Pourquoi dans la technique des **séries appariés**, on utilise la **formule** du test **comparaison de moyennes** ?

Pourquoi on utiliserait pas une **autre formule** spécifique ? »

REPONSE : ON UTILISE UNE FORMULE SPECIFIQUE !

$$\text{si } n > 30 \quad \varepsilon = m_d / \sqrt{\frac{s^2}{n}}$$

$$\text{si } n < 30 \quad t = m_d / \sqrt{\frac{s^2}{n}}$$

m_d étant la moyenne des différences, calculées entre les 2 valeurs du couple

I-6 Réponses aux questions

QUESTION :

« Dans un exemple du cours, on ne suit pas les conditions d'effectif présentes dans le **tableau récapitulatif** pour effectuer les tests de **coefficient de corrélation et r' de Spearman**, pourquoi ? »

I-6 Réponses aux questions

TESTS NON PARAMÉTRIQUES

- Utilisation obligatoire si les effectifs sont trop faibles (< 5), avec des caractères quantitatifs. Dans ce cas, les populations ne se distribuent pas normalement.
- Présentent une excellente robustesse

LIAISON QUANTITATIFS / QUALITATIFS

U de Mann & Whitney

LIAISON ENTRE QUANTITATIFS

Spearman (= Corrélation)

I-6 Réponses aux questions

Sur un échantillon de 10 sujets d'âges différents, on recueille les données suivantes : âge (années) et concentration de cholestérol dans le sang (g/L)

Effectif	Données Quantitatives	Données Qualitatives	Données Qualitatives - Quantitatives
>4 & <12	r' de Spearman	Comp % ou χ^2	U Mann & Withney
>12 & < 30	Coeff de corrélation r	Comp % ou χ^2	t Student
> 30	Coeff de corrélation r	Comp % ou χ^2	Comp moyennes

I-6 Réponses aux questions

QUESTION :

« Dans un exemple du cours, on ne suit pas les conditions d'effectif présentes dans le **tableau récapitulatif** pour effectuer les tests de **coefficient de corrélation et r' de Spearman**, pourquoi ? »

REPONSE : En présence de variables quantitatives, pour des effectifs >4 et <12 , on est en droit de se poser la question. De toutes façons, on ne fait JAMAIS d'étude avec $n < 4$
t Student ou U Mann et Whitney ??

Par ex : $n=10$: t Student ou U Mann et Whitney
 $n=8$: plutôt U Mann et Whitney

I-6 Réponses aux questions

QUESTION :

« Quelle est la **formule** du calcul du **degré de liberté** dans cette même **situation** car elle n'est **pas donnée** dans le cours ? Est ce juste à titre illustratif ou **faut il la savoir** ? »

I-6 Réponses aux questions

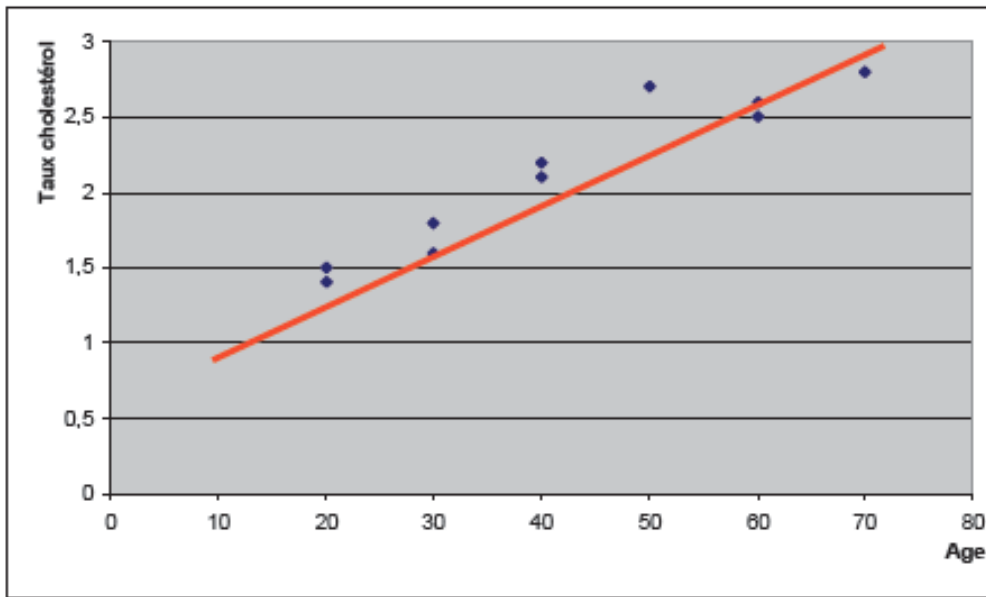
r calculé = 0,955

r théorique ($\alpha = 5\%$) lu dans la table avec $10-2 = 8$ ddl = 0,632

r calculé > r théorique

Rejet de H_0

Il existe une relation significative ($\alpha = 5\%$) entre l'âge et le taux de cholestérol



Plus l'âge augmente,
plus le taux de
cholestérol augmente

I-6 Réponses aux questions

QUESTION :

« Quelle est la **formule** du calcul du **degré de liberté** dans cette même **situation** car elle n'est **pas donnée** dans le cours ? Est ce juste à titre illustratif ou **faut il la savoir** ? »

REPONSE : La formule (non donnée en cours, pardon !!) est **$ddl = n-2$** , n étant le nb de couples (x,y).

Inutile de savoir cela, **pour le concours**, comme les autres formules..

I-6 Réponses aux questions

QUESTION :

« Est ce possible de repréciser les **conditions** d'effectif pour effectuer les tests de **non paramétriques** ? »

I-6 Réponses aux questions

- Utilisat
faibles
- (< 5),
cas, lo
normal

Effectif	Données Quantitatives	Données Qualitatives	Données Qualitatives - Quantitatives
$>4 \text{ \& } <12$	r' de Spearman	Comp % ou χ^2	U Mann & Withney
$>12 \text{ \& } < 30$	Coeff de corrélation r	Comp % ou χ^2	t Student
> 30	Coeff de corrélation r	Comp % ou χ^2	Comp moyennes

I-6 Réponses aux questions

QUESTION :

« Est ce possible de repréciser les **conditions** d'effectif pour effectuer les tests de **non paramétriques** ? »

REPONSE :

Pour le concours retenir les conditions d'effectif présents dans le tableau récapitulatif !

REPRISE DE NOTIONS !!!

*La répétition est à
la base de l'apprentissage...*



II-1 Reprise de notions

NOTION : « nb degrés de liberté » = **nb de valeurs minimales nécessaires afin de pouvoir calculer celle ou celles manquantes, connaissant la somme de toutes les valeurs..**

$$\sum_{i=1}^n x_i = T \quad \text{Si il manque 1 valeurs on peut la retrouver donc } ddl = n-1$$

1 - Test du χ^2

$$\text{Nb ddl} = (\text{nb lig}-1)(\text{nb col}-1)$$

Exemple 2 lignes, 2 colonnes : **ddl = 1**

CAR 1

	A	B	Total
CAR 2 C	n_1	n_2	Tot C
D	n_3	n_4	Tot D
Total	Tot A	Tot B	T

2 - Test t de Student (n_1 ou $n_2 < 30$)

$$\text{Nb ddl} = (n_1-1)+(n_2-1)$$

3 - Coefficient de corrélation

$$\text{Nb ddl} = n-2$$

II-2 Reprise de notions

NOTION : « IC chevauchent ou non ... »

On compare la diminution du taux de cholestérol produite par un nouveau tt T à celle produite par le tt de référence R.

Après tt dans le groupe T (nT=400) IC95% = [1,10 ; 1,20] mT= 1,15 mmol/l

Après tt dans le groupe R (nR=400) IC95% = [1,30 ; 1,50] mR= 1,4 mmol/l

Si les 2 IC sont **disjoints** : on accepte **H1**

Si les 2 IC se **chevauchent** : on conserve **H0**

II-3 Reprise de notions

NOTION : « score utilisé afin de quantifier les états de dépression »

On veut savoir si une nouvelle molécule présente un effet anti dépresseur. Pour cela, on organise un essai portant sur 20 malades dépressifs, répartis en 2 groupes.

Les 20 malades sont répartis par TAS en 2 groupes de 10 sujets, l'un recevant la nouvelle molécule, l'autre recevant un placebo.

On évalue les patients à l'aide d'une échelle numérique de 0 (non déprimé) à 50 (très déprimé). Le groupe témoin reçoit le placebo.

Les patients des 2 groupes sont évalués avant traitement, puis après traitement au bout de 28 jours.

II-3 Reprise de notions

Témoins Traités

J0	J28	J0	J28
34	31	27	22
30	28	32	25
25	26	30	23
27	25	28	26
31	24	25	20
24	25	33	27
28	26	29	21
30	27	31	26
35	32	32	25
26	25	29	23

Afin de déterminer la dépression on aurait pu utiliser une qualification telle que « un peu déprimé, pas trop déprimé, etc.. ». Ces termes sont imprécis et surtout ne permettent pas de comparer objectivement l'état des patients avant et après TT..

Les **scores numériques** sont utilisés afin d'être précis et de pouvoir comparer J0 et J28.

Score d'Apgar, températures corporelles,..

QCM TIME !!!

8 QCMS en 12 minutes, prêt à relever le duel ???





CORRECTION

TIME !!!

La sentence est irrévocable...



OU



Monsieur HEUREUX

Monsieur GRINCHEUX



IV Le grand récap'

-binaire = **sous catégorie de nominale**

-courbe Gauss -> définir **une marge d'erreur** lors de l'estimation de la moyenne

-population de référence -> **parfaitement définie** -> pas besoin d'échantillonner

-précision = notion générale (bonne, mauvaise), mais **quantifiée par un indice i**

-pour les conditions d'effectif des tests -> suivre le tableau **récap'**

-coefficient de corrélation : **ddl = n-2** (pas à retenir concours)

-nb ddl : **nb de valeurs minimales nécessaires pour calculer celle(s) manquante(s), connaissant la somme de toutes les valeurs.**

-2 IC **disjoints** : accepte **H1** ; 2 IC **chevauchés** : conserve **H0**

-score = variable **quantitative** !

C'est OK ?




Le Séance de révision de Biostat' c'est fini.....

Mais RDV le 1^{er} et le 13 décembre...



Vive la
biostat' !



et.....



Finissons la séance

par la danse de la biostat,

ou plutôt...

LE BIOSTAT' STYLE !



Merci à tous pour votre attention !

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

Prochain rendez-vous biostatistiques :

-1^{er} décembre : **CCB**

-13 décembre : **Séance de révision épisode 2**

Tous droits réservés au tutorat niçois. Le tutorat est gratuit.

Toute reproduction ou vente est interdite.