

## TD 5 : ANOVA à un facteur

**Exercice 1** : On souhaite étudier les effets de trois traitements (AX23, BWW9 et Contrôle) sur le rythme cardiaque chez les patients âgés de 50 à 85 ans, souffrant d'une maladie cardiaque appelée *hypertrophie ventriculaire gauche*. Après que le médicament a été administré, le rythme cardiaque après 5 mn est mesuré. Il y a 32 personnes par traitement. Les données se trouvent dans le fichier **rythme\_cardiaque.sta**. Télécharger les données sur le cours **Méthodes statistiques pour l'analyse des données en psychologie** (téléchargeables à l'adresse <http://coursenligne.u-paris10.fr/>). Ouvrir le fichier de données avec STATISTICA, ensuite créer un classeur avec toutes ces données puis activer la feuille de données.

Le fichier de données **rythme\_cardiaque.sta** contient, pour chaque patient, les observations des variables :

- **traitement** : AX23, BWW9, Contrôle ;
- **rythme cardiaque** : score 5 mn après traitement.

**Remarque 1** : Dans cette étude les trois échantillons sont indépendants . Ces trois groupes de patients ont été constitués par tirage au sort.

Nous voulons répondre à la question suivante : *Les traitements ont-ils des effets significativement différents ?*

### 1) Première partie : Formulation du problème

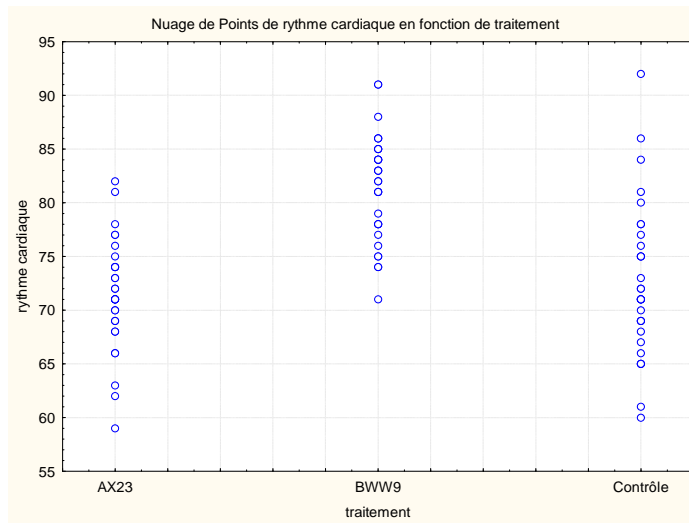
- a) Définir la population étudiée. Identifier la variable dépendante et le facteur ou variable indépendante.
- b) Écrire explicitement les hypothèses nulle et alternative de votre test.
- c) Formuler le modèle théorique de l'ANOVA à un facteur et mentionner les conditions à vérifier.

### 2) Deuxième partie : Analyse descriptive des données

- a) Représenter graphiquement, dans un nuage de points, la variable rythme cardiaque en fonction du traitement.

#### Commandes avec STATISTICA

Utiliser le menu **Graphiques - Nuage de points**. Sélectionner ensuite l'onglet **Base** et indiquer les variables utilisées pour l'analyse (assigner à X : **traitement** et à Y : **rythme cardiaque**). Une fois choisies les variables cliquer sur **OK**. Finalement en bas de la fenêtre dans le **Type d'ajustement** décocher l'item **Linéaire** puis cliquer sur **OK**. On obtient :



b) Moyennes des 3 échantillons et moyenne globale.

- i. Calculer les moyennes observées pour chaque échantillon et donner la valeur de la moyenne globale. Sur la base de ces statistiques descriptives, quelle première conclusion peut-on en tirer ?

### Commandes avec STATISTICA

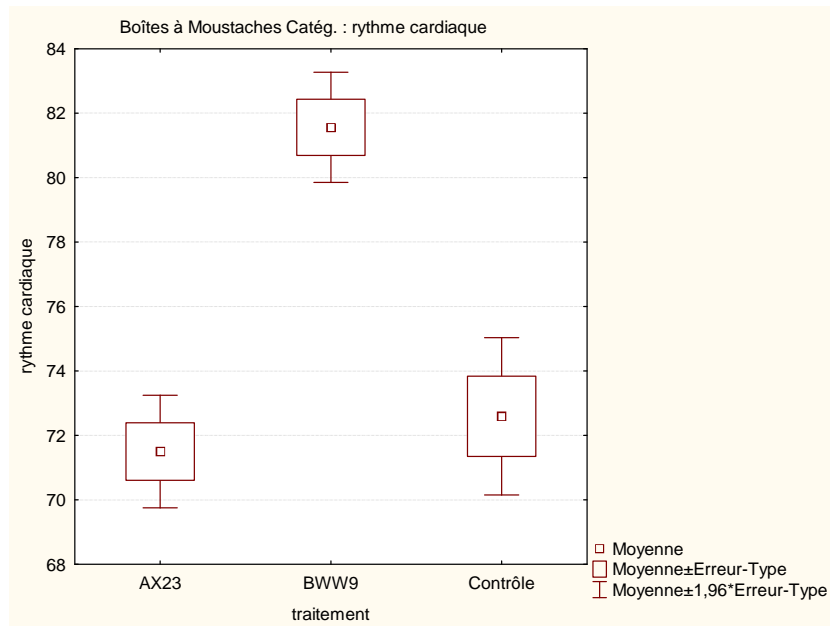
Utiliser le menu **Statistiques - Statistiques élémentaires - Décompositions & ANOVA à un facteur** et cliquer sur **OK**. Maintenant sélectionner l'onglet **ANOVA (tables individuelles)** et indiquer les variables utilisées pour l'analyse (assigner à **variables dépendantes** : `rythme_cardiaque` et à **variables de classement** : `traitement`). Une fois choisies les variables cliquer sur **OK**. Sélectionner ensuite l'onglet **Stats descriptives** et cliquer sur le bouton **Synthèse : Table des statistiques**. On obtient le résultat suivant :

Statistiques Descriptives par Groupes N=96 (aucune VM dans les vars dép.)			
traitement	rythme cardiaque Moyennes	rythme cardiaque N	rythme cardiaque Ec-Type
AX23	71,50000	32	5,041761
BWW9	81,56250	32	4,931384
Contrôle	72,59375	32	7,047577
TsGrpes	75,21875	96	7,278234

- c) Faire des boîtes à moustaches pour regarder visuellement si les moyennes observées sont différentes.

### Commandes avec STATISTICA

Sur le même onglet de **Stats descriptives** (utilisé ci-dessus) cliquer sur le bouton **Boîtes à moustaches catégorisées** et cocher l'item **Moyenne/Erreur-Type/1.96\*Erreur-Type**. On obtient le graphique suivant :



### 3) Troisième partie : Vérification des conditions.

a) Indépendance.

Les tirages effectués pour constituer les échantillons sont-ils aléatoires et indépendants ?

b) Homogénéité des variances ou homoscedasticité.

Donner avec STATISTICA deux méthodes pour vérifier l'homogénéité des variances.

#### Commandes avec STATISTICA

Utiliser le menu Statistiques - Statistiques élémentaires - Décompositions & ANOVA à un facteur, choisir les variables et cliquer sur OK, activer l'onglet ANOVA & tests puis cliquer sur le bouton Tests de Levene ou le bouton Tests de Brown-Forsythe. On obtient ainsi les résultats suivants :

Test de Levene d'Homogénéité des Variances								
Effets significatifs marqués à $p < ,05000$								
Variable	SC Effet	dl Effet	MC Effet	SC Erreur	dl Erreur	MC Erreur	F	p
rythme cardiaque	51,60067	2	25,80033	1198,504	93	12,88714	2,002022	0,140841

Test d'Homogénéité des Variances de Brown-Forsythe								
Effets significatifs marqués à $p < ,05000$								
Variable	SC Effet	dl Effet	MC Effet	SC Erreur	dl Erreur	MC Erreur	F	p
rythme cardiaque	43,18750	2	21,59375	1396,219	93	15,01310	1,438327	0,242554

c) Normalité de chaque variable.

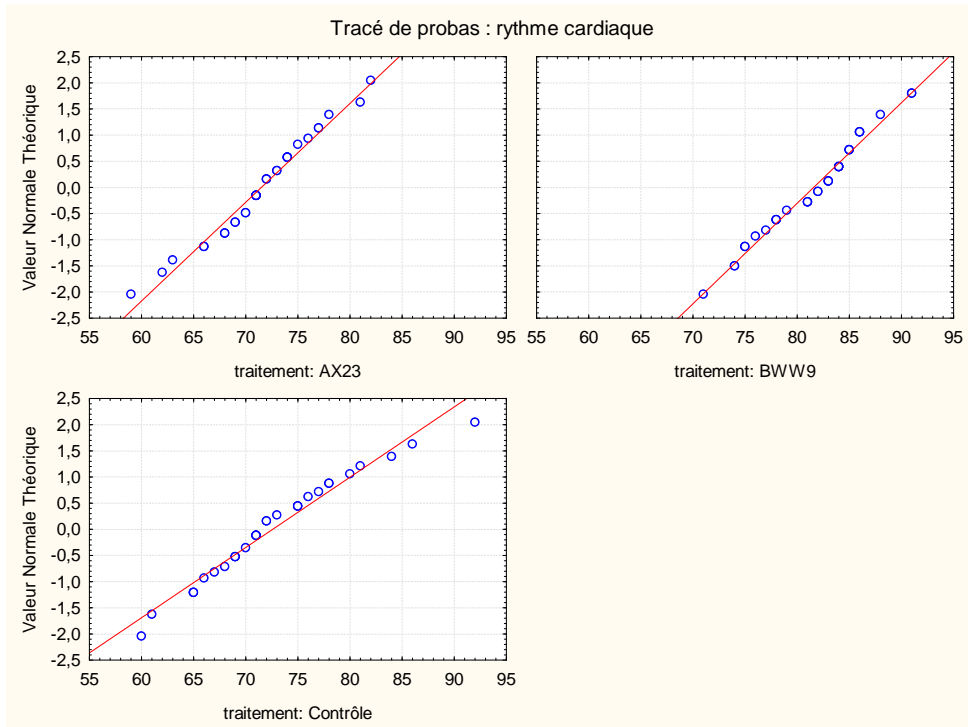
Faire des graphiques de *droite de Henry catégorisée* pour regarder "visuellement" si la distribution de chaque variable  $Y_j$  est normale, de moyenne  $\mu_j$  et de variance  $\sigma^2$ . Sur la base de ces graphiques quelle première conclusion peut-on en tirer ? Faire maintenant des Histogrammes

catégorisés avec deux tests de normalité. Que peut-on en conclure ?

— Droites de Henry catégorisées.

### Commandes avec STATISTICA

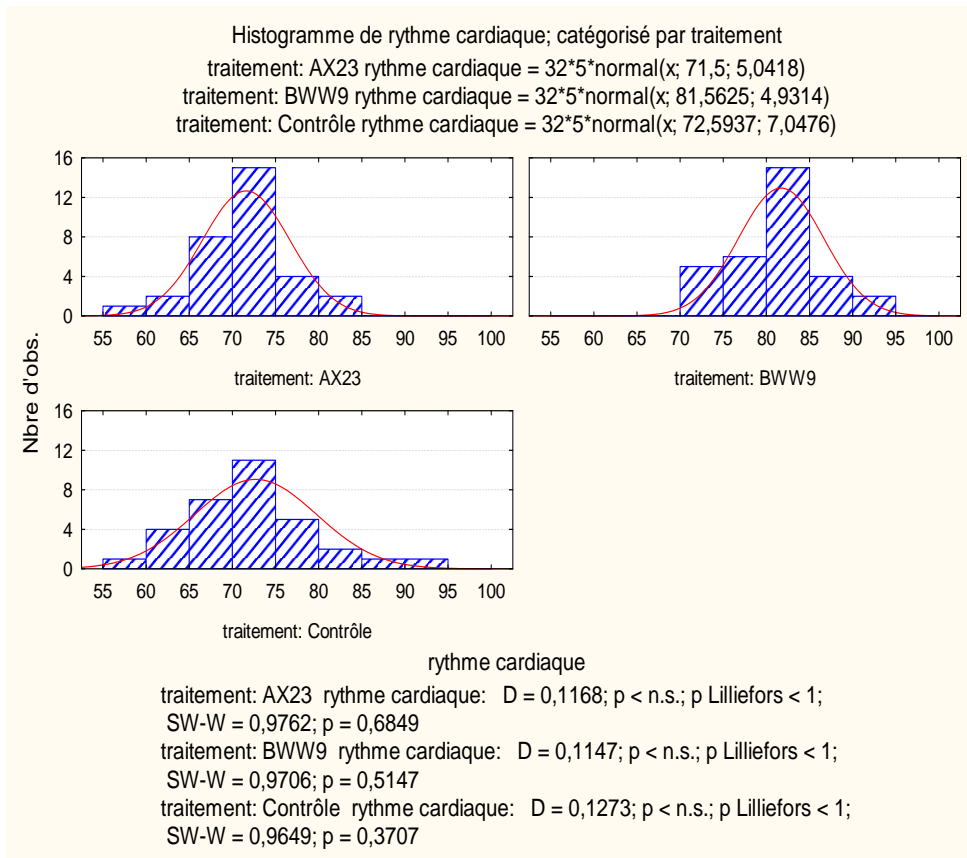
Sur le même onglet de ANOVA & test (utiliser ci-dessus, voir les commandes de la question 3 b ) cliquer sur le bouton Droites de Henry catégorisées. On obtient :



— Histogrammes Catégorisés et test de normalité.

### Commandes avec STATISTICA

Sélectionner dans le menu la commande Graphiques - Histogrammes. Dans l'onglet Base sélectionner la variable rythme cardiaque, puis dans l'onglet Avancé cocher les options Test de Shapiro-Wilk et Test de Kolmogorov-Smirnov. Enfin dans l'onglet Catégorisé cocher Oui dans Catégories de X puis cocher l'item Codes. Cliquer ensuite sur le bouton Spécifier les Codes et sélectionner la variable traitement. Dans la nouvelle fenêtre Codes des Catégories : taper AX23 BWW9 Contrôle puis cliquer sur OK. Finalement, cliquer sur OK pour afficher le graphique ci-dessous :



#### 4) Quatrième partie : Application de l'ANOVA à un facteur.

Variabilités inter-groupes et intra-groupes :

- (a) Donner les *sommes de carrés* (notée SC dans STATISTICA) *inter-groupes* et *intra-groupes*.
- (b) Donner les *variabilités* ou *carrés moyens* (notée MC dans STATISTICA) *inter-groupes* et *intra-groupes*.
- (c) Donner la valeur de la statistique de Fisher F et le nombre de degrés de liberté de la statistique F. Donner l'expression de la *p-valeur* en termes de la valeur de la statistique F (qu'on note dans le cours  $\alpha_{\text{obs}}$ ). Puis donner la valeur de la *p-valeur* fournie par STATISTICA. Que peut-on en conclure au risque  $\alpha = 5\%$ , en précisant le risque associé à la décision prise ?
- (d) Est-ce que le résultat du test montre une différence significative entre les moyennes, au risque 5% ? Cette décision correspond t-elle à votre impression en voyant les graphiques ?

#### Commandes avec STATISTICA

Utiliser le menu Statistiques - ANOVA et sélectionner ensuite dans l'onglet Base l'option ANOVA à un facteur et Spécifications rapides, cliquer ensuite sur le bouton OK puis sur le bouton à droite Tous les effets. On obtient le résultat suivant :

Effet	Tests Univariés de Significativité pour rythme cardiaque Paramétrisation sigma-restreinte Décomposition efficace de l'hypothèse				
	SC	Degr. de Liberté	MC	F	p
ord. origine	543154,6	1	543154,6	16391,97	0,000000
traitement	1950,8	2	975,4	29,44	0,000000
Erreur	3081,6	93	33,1		

5) **Cinquième partie : Test post hoc après une ANOVA à un facteur.**

L'ANOVA précédente permet de conclure qu'il existe au moins une différence significative entre les moyennes, mais n'indique pas quelles sont les paires d'échantillons pour lesquelles ces différences de moyennes sont significatives. Différents tests, appelés *Test post hoc*, ont été proposés pour étudier cette question.

a) Test LSD (least significant difference) de Fisher.

Commandes avec STATISTICA

Reprenons le menu Statistiques - Statistiques élémentaires - Décompositions & ANOVA à un facteur en indiquant comme précédemment (voir première, deuxième partie ou troisième partie ) la variable dépendante et la variable de classement. Une fois définies les variables cliquer sur le bouton OK. Utiliser l'onglet ANOVA (tables individuelles) puis l'onglet Tests post-hoc. Cliquer sur le bouton Test LSD ou comparaisons planifiées. On obtient le tableau suivant :

Test LSD ; variable rythme cardiaque Probabilités des Tests Post Hoc Erreur : MC Inter = 33,135, dl = 93,000				
Cellule N°	traitement	{1}	{2}	{3}
		71,500	81,563	72,594
1	AX23		0,000000	0,449158
2	BWW9	0,000000		0,000000
3	Contrôle	0,449158	0,000000	

b) Test de Bonferroni, test de Newman-Keuls, test de Duncan, test HSD de Tukey, test de Scheffé.

Commandes avec STATISTICA

Répéter l'instruction précédente (voir commandes avec STATISTICA, question 5a) mais juste à la fin cliquer sur le bouton du test à faire. On obtient les tableaux suivants :

Test de Bonferroni ; variable rythme cardiaque				
Probabilités des Tests Post Hoc				
Erreur : MC Inter = 33,135, dl = 93,000				
Cellule N°	traitement	{1}	{2}	{3}
		71,500	81,563	72,594
1	AX23		0,000000	1,000000
2	BWW9	0,000000		0,000000
3	Contrôle	1,000000	0,000000	

Test de Newman-Keuls ; variable rythme cardiaque				
Probabilités Approximatives des Tests Post Hoc				
Erreur : MC Inter = 33,135, dl = 93,000				
Cellule N°	traitement	{1}	{2}	{3}
		71,500	81,563	72,594
1	AX23		0,000106	0,449291
2	BWW9	0,000106		0,000112
3	Contrôle	0,449291	0,000112	

Test de Duncan ; variable rythme cardiaque				
Probabilités Approximatives des Tests Post Hoc				
Erreur : MC Inter = 33,135, dl = 93,000				
Cellule N°	traitement	{1}	{2}	{3}
		71,500	81,563	72,594
1	AX23		0,000053	0,449291
2	BWW9	0,000053		0,000112
3	Contrôle	0,449291	0,000112	

Test HSD de Tukey ; variable rythme cardiaque				
Probabilités Approximatives des Tests Post Hoc				
Erreur : MC Inter = 33,135, dl = 93,000				
Cellule N°	traitement	{1}	{2}	{3}
		71,500	81,563	72,594
1	AX23		0,000106	0,728395
2	BWW9	0,000106		0,000106
3	Contrôle	0,728395	0,000106	

Test de Scheffé ; variable rythme cardiaque				
Probabilités des Tests Post Hoc				
Erreur : MC Inter = 33,135, dl = 93,000				
Cellule N°	traitement	{1}	{2}	{3}
		71,500	81,563	72,594
1	AX23		0,000000	0,749813
2	BWW9	0,000000		0,000000
3	Contrôle	0,749813	0,000000	

6) **Sixième partie : Paramètres du modèle, coefficient  $R^2$  et normalité de résidus.**

- a) Donner la valeur de  $R^2$  fournie par STATISTICA. Quelle conclusion peut-on en tirer ? Donner les valeurs observées de MC Modèle, SC Modèle, MC Résidu et SC Résidu ? Que peut-on remarquer ?

Commandes avec STATISTICA

Utiliser le menu Statistiques - ANOVA et sélectionner ensuite dans l'onglet Base l'option ANOVA à un facteur et Spécifications rapides puis cliquer ensuite sur le bouton OK. Ajouter la variable dépendante et le facteur puis cliquer sur OK. En bas de la fenêtre d'ANOVA cliquer sur l'onglet Autres résultats. Enfin sélectionner l'onglet Synthèse et cliquer sur le bouton R modèle complet. On obtient le tableau suivant :

Dépendnt Variable	Test de la SC du modèle entier vs. SC Résiduels							
	Multiple R	Multiple R <sup>2</sup>	Ajusté R <sup>2</sup>	SC Modèle	dl Modèle	MC Modèle	SC Résidu	dl Résidu
rythme cardiaque	0,622615	0,387650	0,374481	1950,813	2	975,4063	3081,594	93

**Remarque 2 :** Dans le tableau ci-dessus nous avons arrondi les valeurs à trois décimales pour diminuer la largeur du tableau.

- b) Donner les coefficients du modèle.

Pour calculer les coefficients du modèle répéter l'instruction ci-dessus : Commandes avec STATISTICA de la question 6 a. On obtient le tableau ci-dessous :

Effet	Estimations de Paramètres Paramétrisation sigma-restreinte							
	Niveau Effet	Colonne	rythme cardiaque Param.	rythme cardiaque Err-Ty.	rythme cardiaque t	rythme cardiaque p	-95,00% Lmt Cnf.	+95,00% Lmt Cnf.
ord. origine		1	75,21875	0,587504	128,0311	0,000000	74,05208	76,38542
traitement	AX23	2	-3,71875	0,830856	-4,4758	0,000022	-5,36866	-2,06884
traitement	BWW9	3	6,34375	0,830856	7,6352	0,000000	4,69384	7,99366

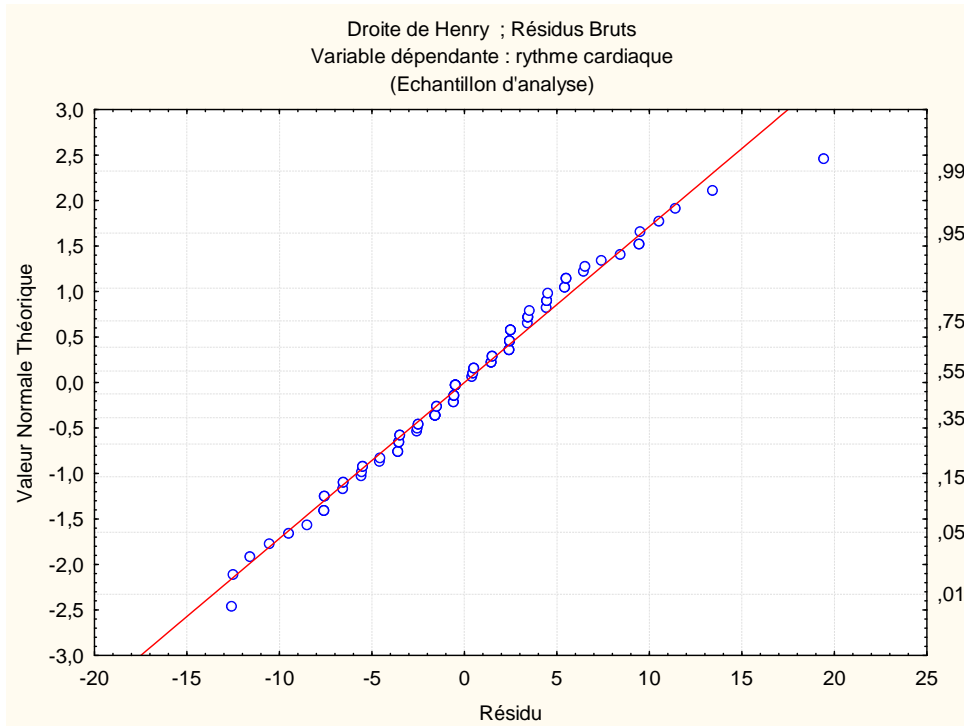
- c) Regarder la normalité des résidus à l'aide de la droite de Henry des résidus. Donner les valeurs prévues et résidus. Que peut-on remarquer sur les valeurs prévues ? Comment s'obtiennent ces valeurs ? (suggestion : regarder les coefficients du modèle). Comment s'obtiennent les résidus ? (suggestion : regarder les valeurs observées et les valeurs prévues de la variable rythme cardiaque).

Commandes avec STATISTICA

Utiliser le menu Statistiques - ANOVA et sélectionner ensuite dans l'onglet Base l'option ANOVA à un facteur et Spécifications rapides puis cliquer ensuite sur le bouton OK. Ajouter la variable



dépendante et le facteur puis cliquer sur OK. En bas de la fenêtre d'ANOVA cliquer sur l'onglet Autres résultats. Enfin sélectionner l'onglet Résidus 1 et cliquer par exemple sur le bouton Droite de Henry. On obtient le graphique suivant :



Pour calculer les valeurs prévues et résidus du modèle répéter l'instruction ci-dessus : Commandes avec STATISTICA de la question 6 b. On obtient le tableau suivant :

Numéro d'obs.	Valeurs Observées, Prévues, et Résidus Paramétrisation sigma-restreinte (Echantillon d'analyse)		
	rythme cardiaque Observées	rythme cardiaque Prévues	rythme cardiaque Résids
1	72,00000	71,50000	0,5000
2	78,00000	71,50000	6,5000
3	71,00000	71,50000	-0,5000
4	72,00000	71,50000	0,5000
5	66,00000	71,50000	-5,5000
6	74,00000	71,50000	2,5000
7	62,00000	71,50000	-9,5000
8	69,00000	71,50000	-2,5000
9	71,00000	71,50000	-0,5000
10	74,00000	71,50000	2,5000
11	59,00000	71,50000	-12,5000

**Exercice 2 :** On souhaite comparer l'efficacité de 4 dentifrices chez les patients âgés de 15 à 45 ans. Chacun des 4 dentifrices (DentifriceT1, DentifriceT2, DentifriceT3, DentifriceT4) a été testé sur 50 personnes afin que soit mesurée leur impact sur la blancheur des dents. Après que le dentifrice a été administré, la blancheur des dents est mesurée. Les 4 groupes de patients ont été constitués par tirage au sort. De plus, dans cette étude les 4 échantillons sont indépendants.

Les données se trouvent dans le fichier **dentidrice.sta**. Télécharger les données sur le cours **Méthodes statistiques pour l'analyse des données en psychologie** (téléchargeables à l'adresse <http://coursenligne.u-paris10.fr/>). Ouvrir les données avec STATISTICA, ensuite créer un classeur avec toutes ces données puis activer la fenêtre. Ce fichier contient, pour chaque patient, les observations des variables : **traitement** et **blancheur**. Nous voulons répondre à la question suivante : *les dentifrices ont-ils des effets significativement différents ?*

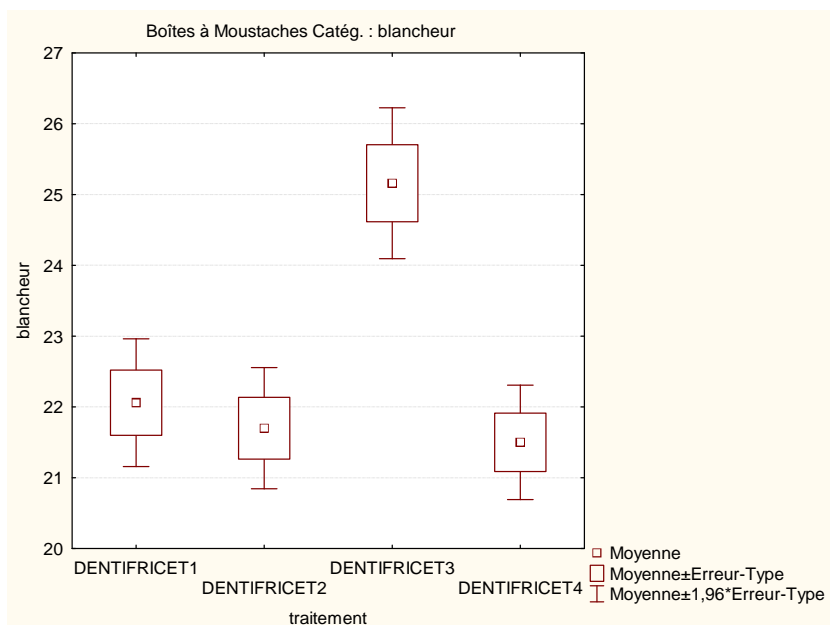
## A vous de jouer !

### 1) Première partie : Formulation du problème

- Définir la population étudiée. Identifier la variable dépendante et le facteur ou variable indépendante.
- Écrire explicitement les hypothèses nulle et alternative de votre test.
- Formuler le modèle théorique de l'ANOVA à un facteur et mentionner les conditions à vérifier.

### 2) Deuxième partie : Analyse descriptive des données

Faire seulement des boîtes à moustaches pour regarder "visuellement" si les moyennes observées sont différentes.



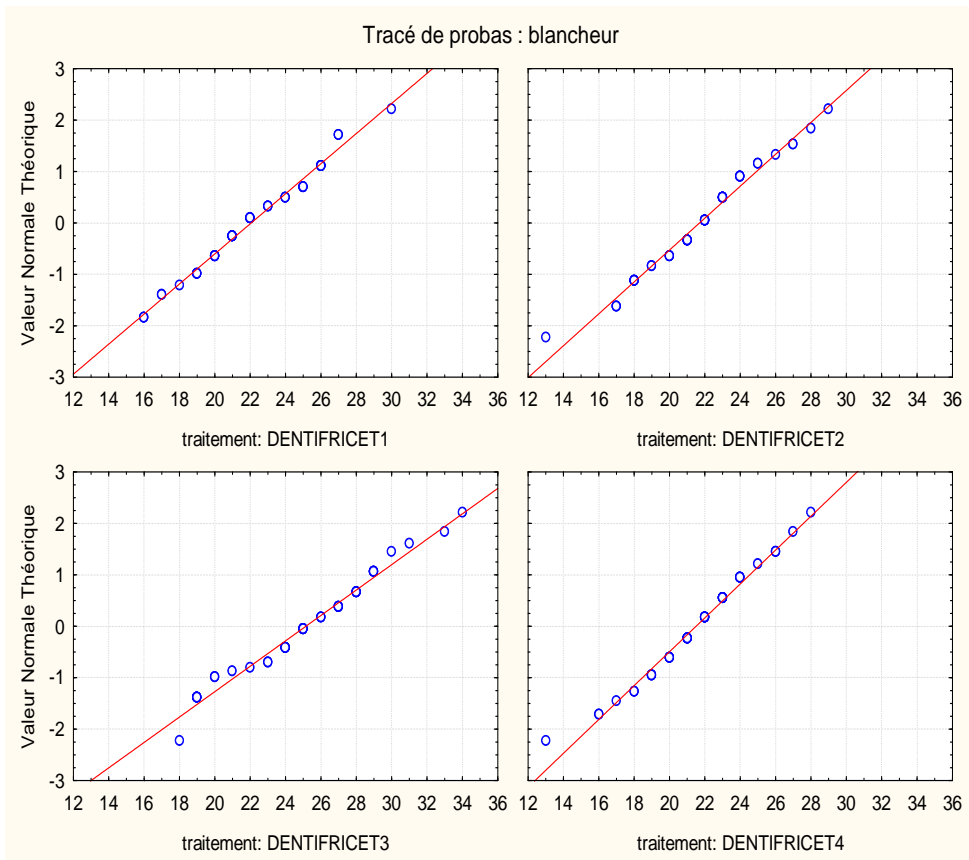
### 3) Troisième partie : Vérification de conditions.

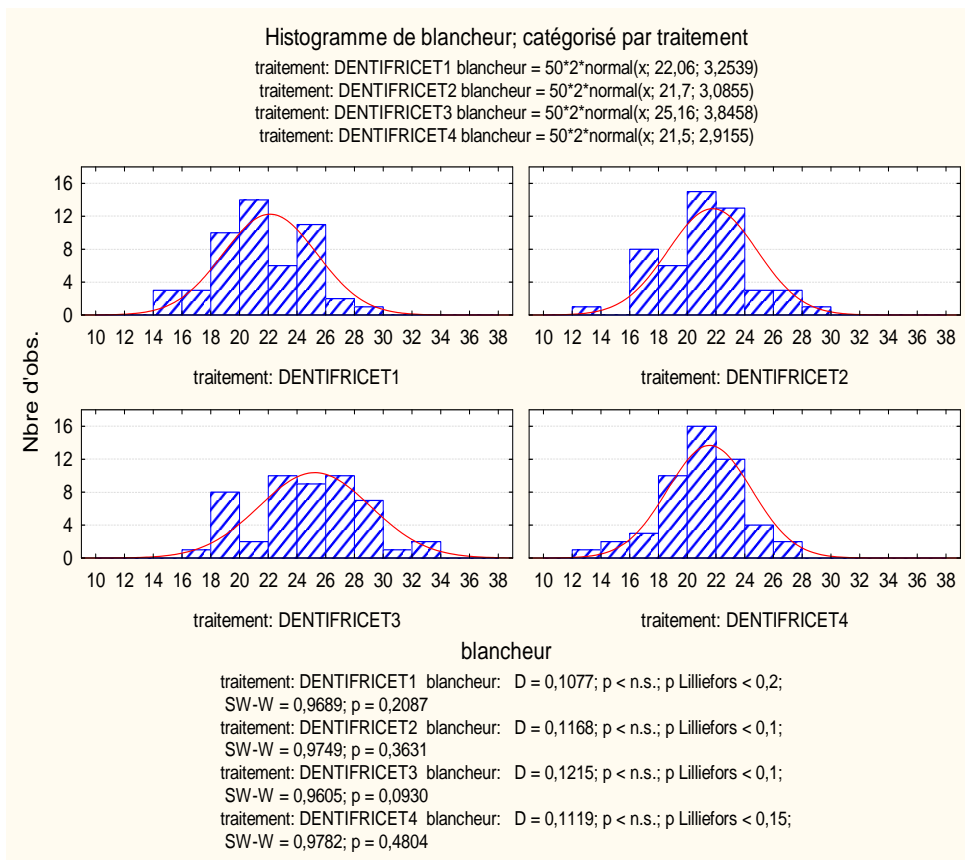
- Indépendance : Cette condition est déjà vérifiée.

- b) Homogénéité des variances.
- c) Normalité de chaque variable.

Test de Levene d'Homogénéité des Variances								
Effets significatifs marqués à $p < ,05000$								
Variable	SC Effet	dl Effet	MC Effet	SC Erreur	dl Erreur	MC Erreur	F	p
blancheur	20,69058	3	6,896861	794,9246	196	4,055738	1,700520	0,168238

Test d'Homogénéité des Variances de Brown-Forsythe								
Effets significatifs marqués à $p < ,05000$								
Variable	SC Effet	dl Effet	MC Effet	SC Erreur	dl Erreur	MC Erreur	F	p
blancheur	20,81500	3	6,938333	816,2800	196	4,164694	1,665989	0,175656





4) **Quatrième partie : Application de l'ANOVA à un facteur.**

- a) Donner les *sommes de carrés inter-groupes* et *intra-groupes*.
- b) Donner les *variabilités inter-groupes* et *intra-groupes*.
- c) Est-ce que le résultat du test montre une différence significative entre les moyennes, au risque 5%? S'il existe une différence entre les 4 dentifrices, appliquer seulement le test *post hoc* de Tukey et de Scheffé.

Effet	Tests Univariés de Significativité pour blancheur Paramétrisation sigma-restreinte Décomposition efficace de l'hypothèse				
	SC	Degr. de Liberté	MC	F	p
ord. origine	102197,2	1	102197,2	9419,363	0,000000
traitement	443,3	3	147,8	13,618	0,000000
Erreur	2126,5	196	10,8		

Dépendnt Variable	Test de la SC du modèle entier vs. SC Résiduels								
	Multiple R	Multiple R <sup>2</sup>	Ajusté R <sup>2</sup>	SC Modèle	dl Modèle	MC Modèle	SC Résidu	dl Résidu	MC Résidu
blancheur	0,415315	0,172487	0,159820	443,2550	3	147,7517	2126,540	196	10,84969

5) Cinquième partie : Test post hoc après une ANOVA à un facteur.

Test HSD de Tukey ; variable blancheur Probabilités Approximatives des Tests Post Hoc Erreur : MC Inter = 10,850, dl = 196,00					
Cellule N°	traitement	{1}	{2}	{3}	{4}
		22,060	21,700	25,160	21,500
1	DENTIFRICET1		0,947533	0,000022	0,830413
2	DENTIFRICET2	0,947533		0,000008	0,990293
3	DENTIFRICET3	0,000022	0,000008		0,000008
4	DENTIFRICET4	0,830413	0,990293	0,000008	

Test de Scheffé ; variable blancheur Probabilités des Tests Post Hoc Erreur : MC Inter = 10,850, dl = 196,00					
Cellule N°	traitement	{1}	{2}	{3}	{4}
		22,060	21,700	25,160	21,500
1	DENTIFRICET1		0,960187	0,000104	0,867756
2	DENTIFRICET2	0,960187		0,000010	0,992736
3	DENTIFRICET3	0,000104	0,000010		0,000003
4	DENTIFRICET4	0,867756	0,992736	0,000003	

6) Sixième partie : Coefficient  $R^2$ , paramètres du modèle, et normalité de résidus.

Test de la SC du modèle entier vs. SC Résiduels									
Dépendnt Variable	Multiple R	Multiple R <sup>2</sup>	Ajusté R <sup>2</sup>	SC Modèle	dl Modèle	MC Modèle	SC Résidu	dl Résidu	MC Résidu
blancheur	0,415315	0,172487	0,159820	443,2550	3	147,7517	2126,540	196	10,84969

Estimations de Paramètres Paramétrisation sigma-restreinte							
Effet	Niveau Effet	Colonne	blancheur Param.	blancheur Err-Ty.	blancheur t	blancheur p	-95,00% Lmt Cnf.
ord. origine		1	22,60500	0,232913	97,05340	0,000000	22,14566
traitement	DENTIFRICET1	2	-0,54500	0,403417	-1,35096	0,178266	-1,34060
traitement	DENTIFRICET2	3	-0,90500	0,403417	-2,24334	0,025994	-1,70060
traitement	DENTIFRICET3	4	2,55500	0,403417	6,33339	0,000000	1,75940

