

# **Relation entre deux variables continues = analyse de corrélation**

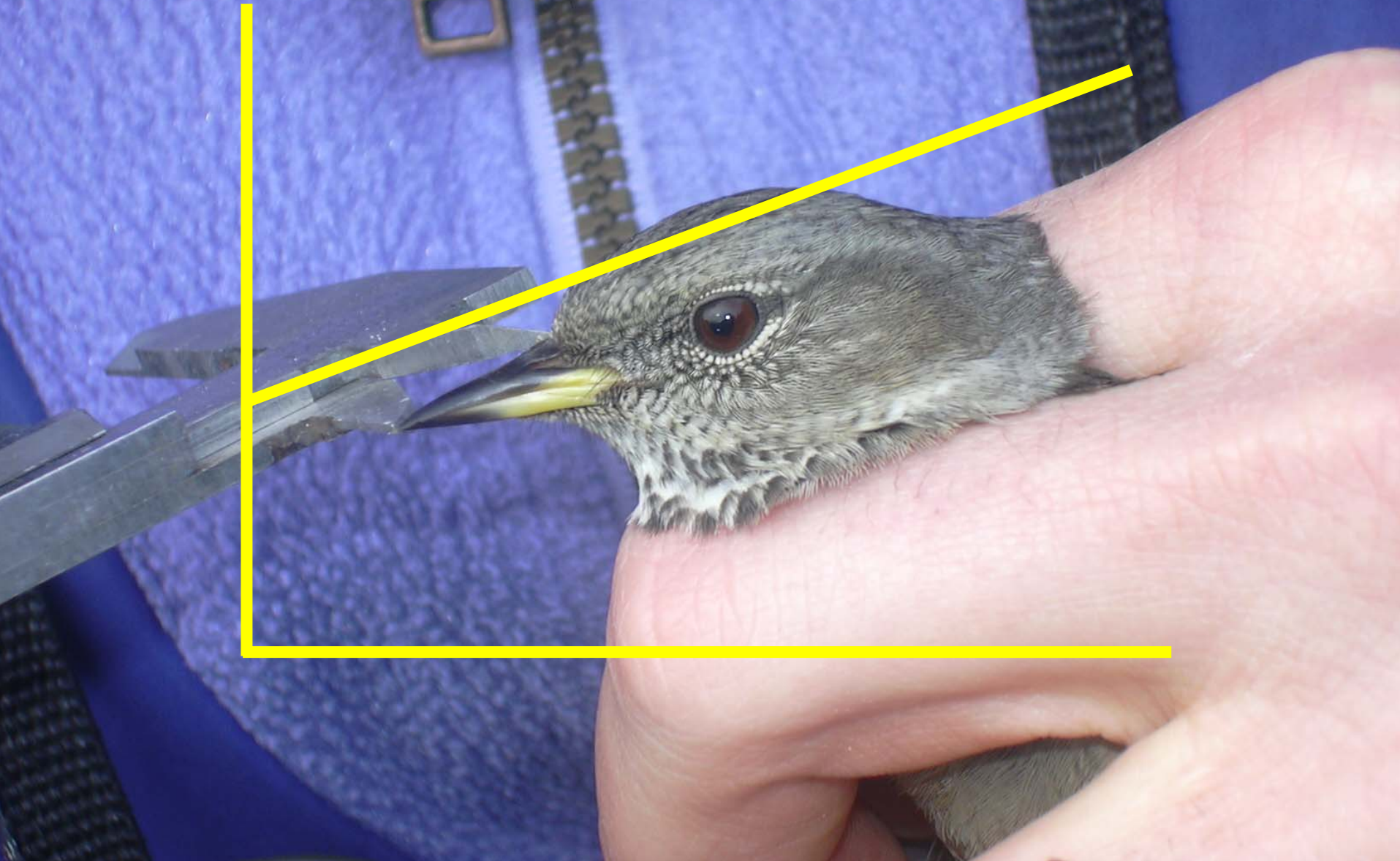
Pierre-Yves Henry  
henry@mnhn.fr



# Quelle analyse ?

Y \ X	Var. catégorielle	Var. continue
	Var. catégorielle	Var. continue
Var. catégorielle	Analyse de fréquence	
Var. continue	Comparaison de moyenne ANOVA	Corrélation Régression

# Test de corrélation





# Test de corrélation

## Intensité de la relation entre deux variables continues

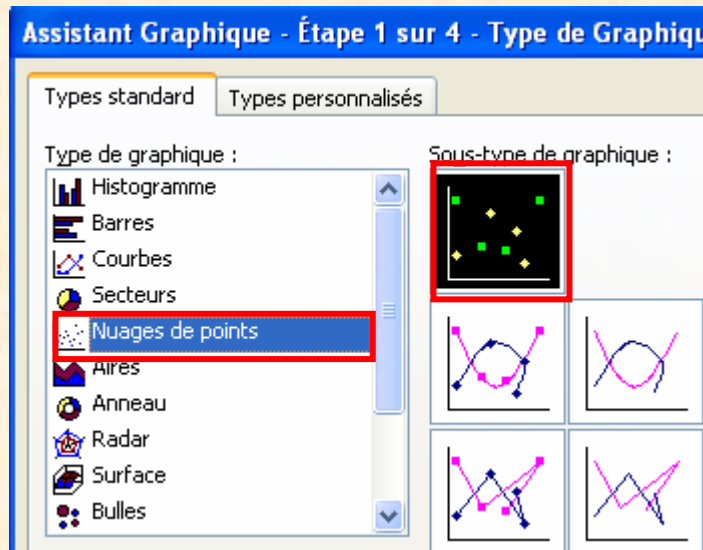
Ex: relation entre masse et longueur d'aile et (ERIRUB, STOC)

Etape 1: visualisation graphique de la relation entre les deux variables par nuage de points

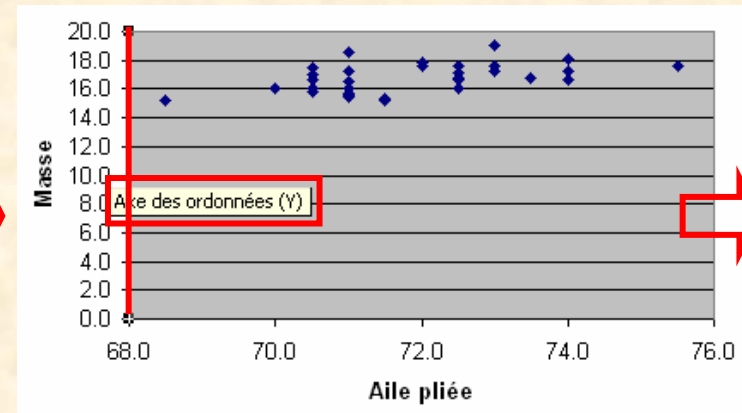
Sélectionner les données

	A	B
1	P	MA
2	68.5	15.2
3	70.0	16.0
4	70.5	17.5
5	70.5	16.7
6	70.5	17.0
7	70.5	15.8
8	70.5	15.8
9	70.5	16.1
10	71.0	17.3
11	71.0	18.6
12	71.0	16.0
13	71.0	15.7
14	71.0	15.5
15	71.0	15.6
16	71.0	16.5
17	71.5	15.3
18	71.5	15.2
19	72.0	17.6
20	72.0	17.8

Sélectionner le type de graphique



Ajuster les axes du graphique en double-cliquant sur l'axe



# Test de corrélation

## Intensité de la relation entre deux variables continues

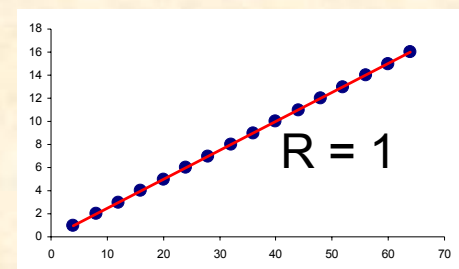
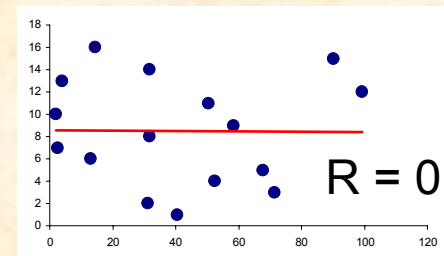
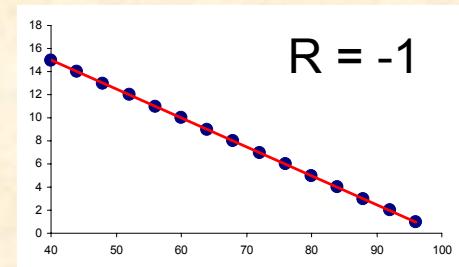
But: quantifier l'intensité de la relation -> coefficient de corrélation

Valeur variant entre:

-1 → corrélation négative parfaite

0 → absence de corrélation

1 → corrélation positive parfaite



# Test de corrélation

Etape 1: visualisation graphique de la relation entre les deux variables par nuage de points

**Format de l'axe**

Motifs Échelle Police Nombre Alignement

Échelle de l'axe des ordonnées (Y)

Automatique

☐ Minimum : 14

☒ Maximum : 20

☒ Unité principale : 2

☒ Unité secondaire : 0,4

☒ Axe des ordonnées (X) coupe à : 0

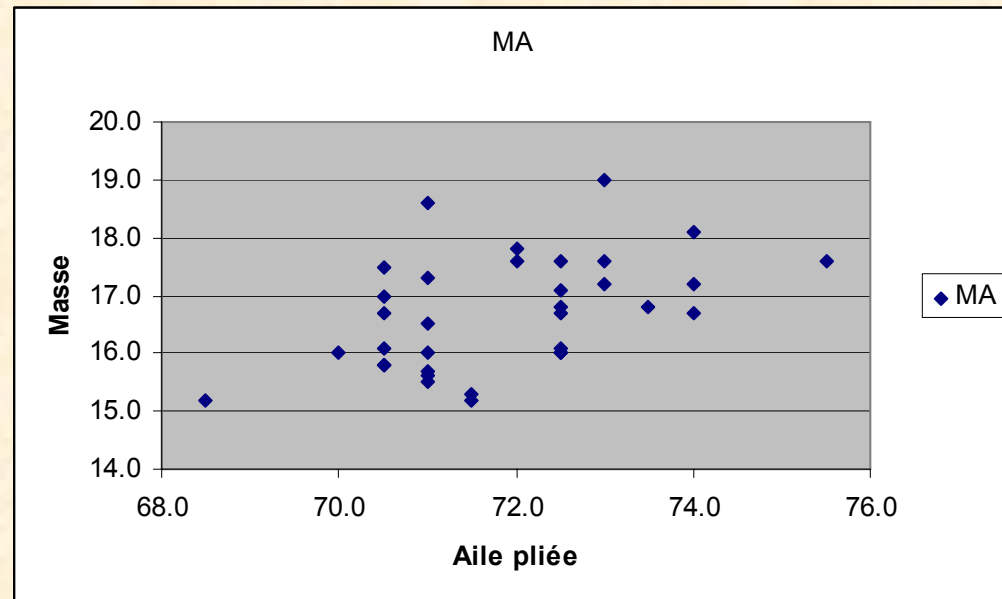
Unités d'affichage : Aucune ☐ Afficher les unités su

☐ Échelle logarithmique

☐ Valeurs en ordre inverse

☐ Axe des ordonnées(X) coupe à la valeur maximale

OK

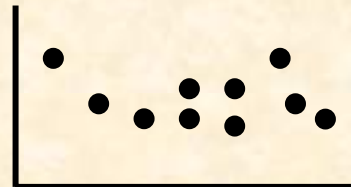


# Test de corrélation

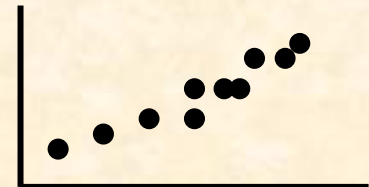
Conditions d'application:  
Homoscédasticité  
Variable normale

Etape 1: visualisation graphique de la relation entre les deux variables par nuage de points. **Que regarder sur le graphique ?**

(1) présence d'une relation ou pas ?

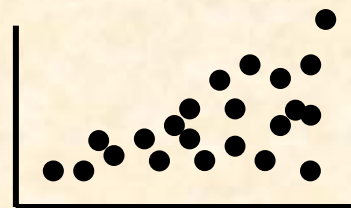


Non ?

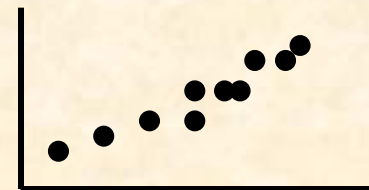


Oui ?

(2) hypothèse de variance homogène ?



Non ?



Oui ?

Si non, essayer de transformer les variables





# Test de corrélation

Etape 2: calcul de la **statistique** = le **coefficient de corrélation de Pearson**  
=COEFFICIENT.CORRELATION(colonne1;colonne2)

Etape 3: Nb de **ddl** = nb de lignes de données - 2

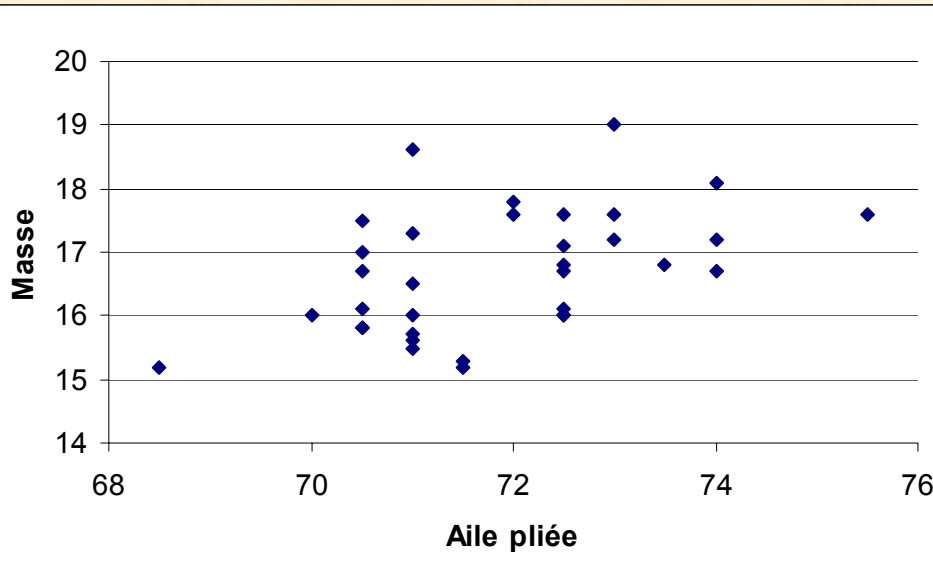
Etape 4: **Probabilité** que la valeur du coefficient de corrélation ne soit pas  
différente de 0 (absence de corrélation) = voir abaque

Nb ddl	$P \leq 0.05$	$P \leq 0.01$
3	0.878	0.959
4	0.811	0.917
5	0.754	0.874
6	0.707	0.834
7	0.666	0.798
8	0.632	0.765
9	0.602	0.735
10	0.576	0.708
15	0.482	0.606
20	0.423	0.537
30	0.349	0.449
50	0.273	0.354
100	0.195	0.254



# Test de corrélation

Ex. Intensité de la relation entre masse et longueur d'aile chez ERIRUB



(2) Calcul du coefficient de corrélation de Pearson

Coeff. corrél = 0.481

Nb ddl 32 (nb d'individus mesurés - 2)

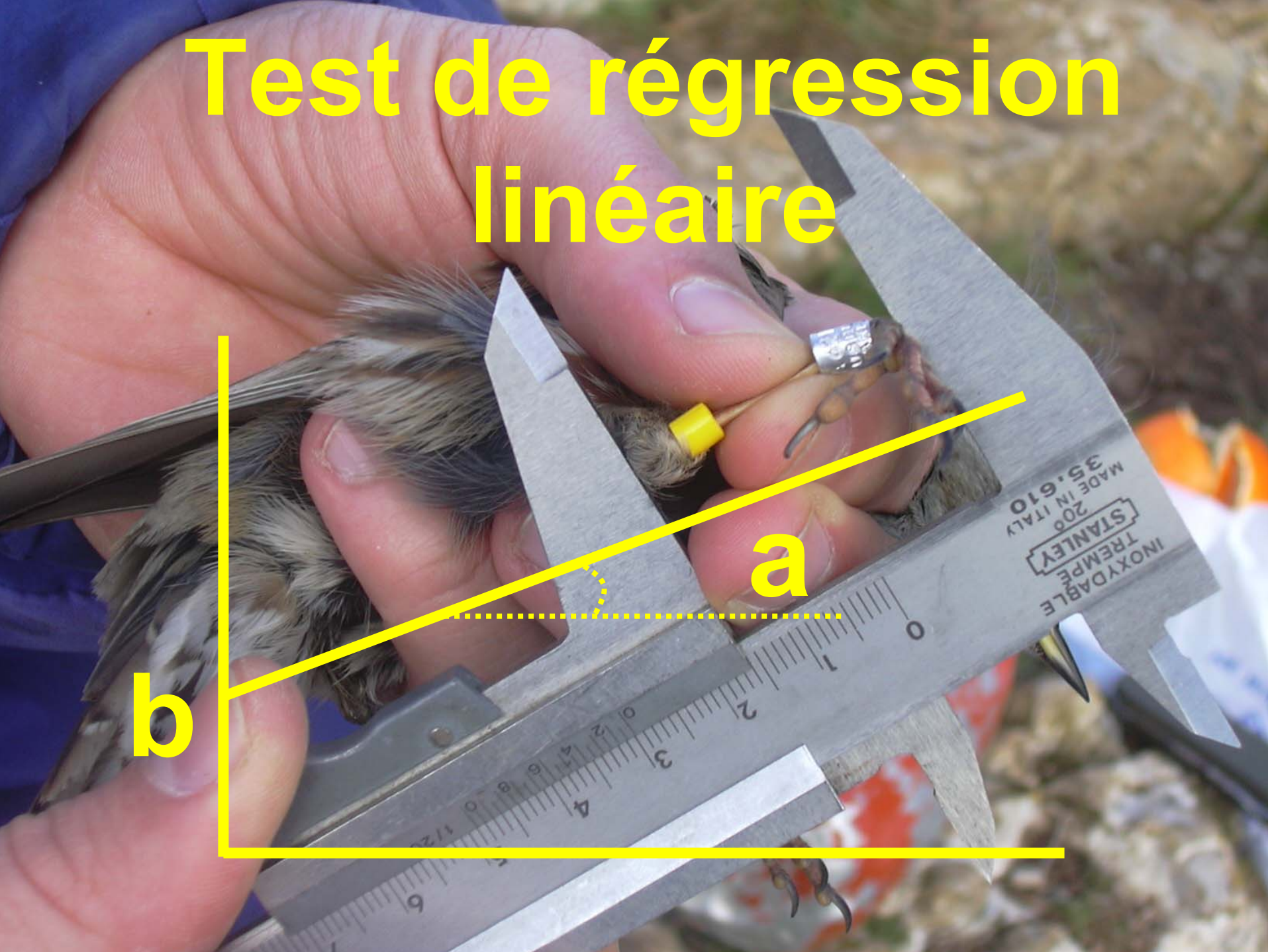
P Identifier le seuil à partir de l'abaque

Abaque pour significativité du coefficient de corrélation

ddl	$P \leq 0.05$	$P \leq 0.01$
3	0.878	0.959
4	0.811	0.917
5	0.754	0.874
6	0.707	0.834
7	0.666	0.798
8	0.632	0.765
9	0.602	0.735
10	0.576	0.708
15	0.482	0.606
20	0.423	0.537
30	0.349	0.449
50	0.273	0.354
100	0.195	0.254

Il y a une corrélation entre masse et taille ( $\rho = 0.481$ ,  $ddl = 32$ ,  $P < 0.05$ )

# Test de régression linéaire





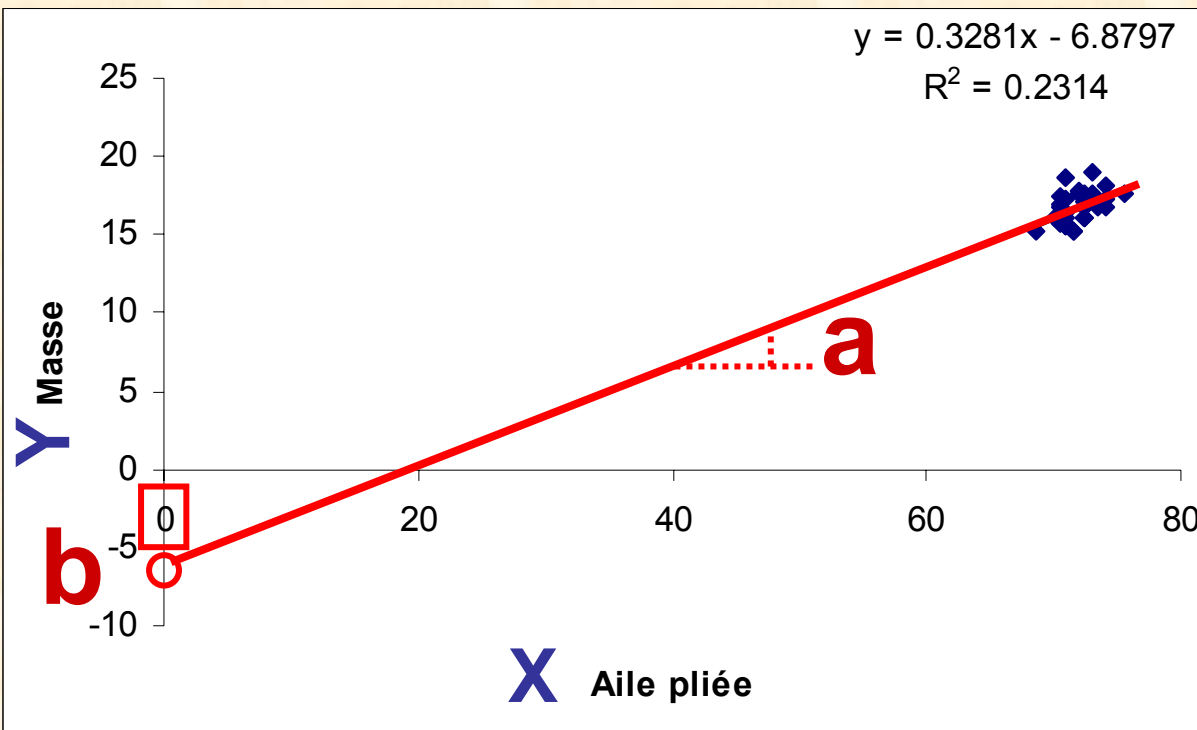
# Test de régression linéaire

## Qu'implique le test d'une régression linéaire ?

-> il existe une relation entre les deux variables

-> cette relation est linéaire et peut être résumé par une fonction mathématique du type:

$$Y = a \times X + b$$



**a** = pente ou coefficient directeur

*Variation de Y pour 1 unité de X*

**b** = ordonnée à l'origine

*Valeur de Y pour X = 0*

Dans ce cas:

$$\text{Masse} = a \times \text{LP} + b$$



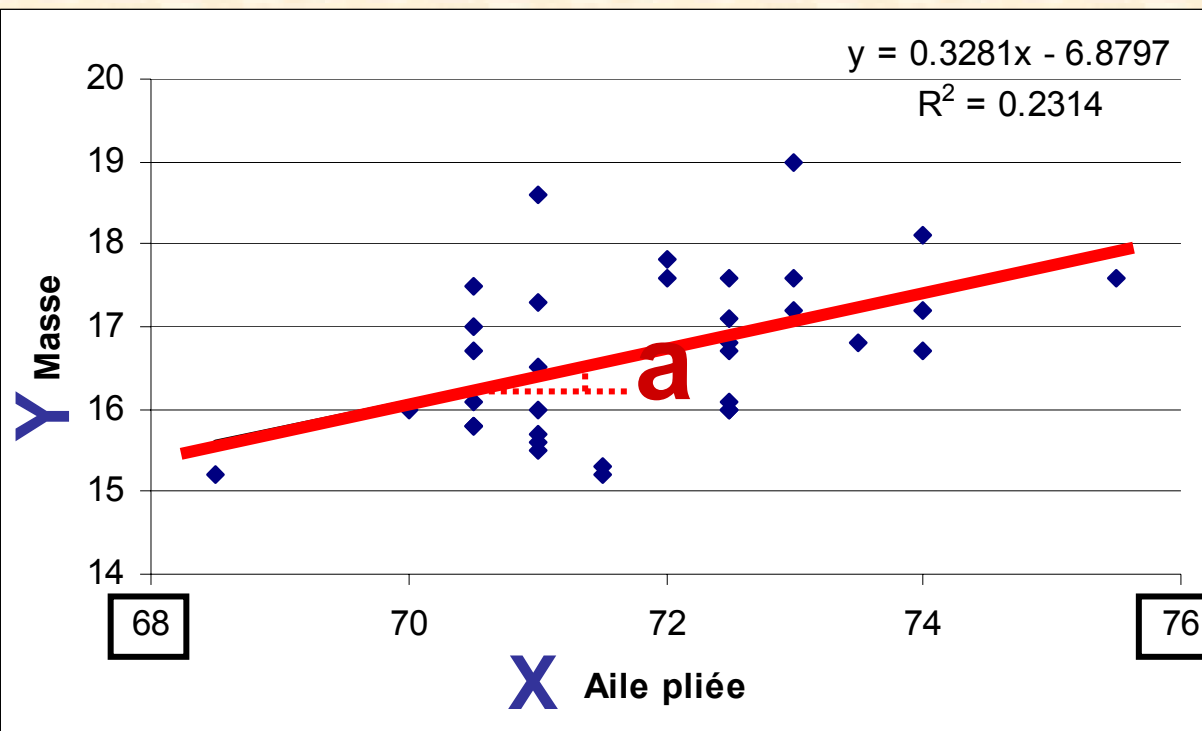
# Test de régression linéaire

## Qu'implique le test d'une régression linéaire ?

-> il existe une relation entre les deux variables

-> cette relation est linéaire et peut être résumé par une fonction mathématique du type:

$$Y = a \times X + b$$



**a = pente ou coefficient directeur**

**Variation de Y pour 1 unité de X**

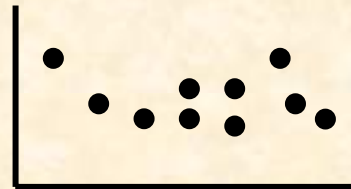
# Test de régression linéaire

## Relation numérique deux variables continues: la dépendance de Y envers X

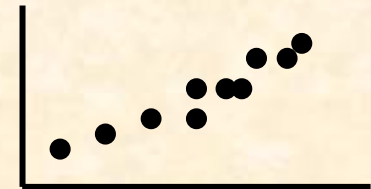
Etape 1: visualisation graphique de la relation entre les deux variables  
par nuage de points

Que regarder sur le graphique ?

(1) présence d'une relation ou pas ?



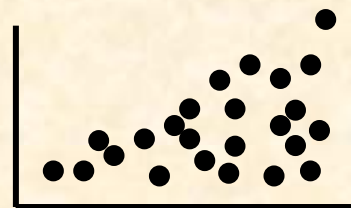
Non ?



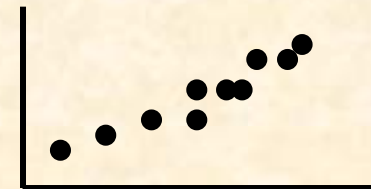
Oui ?

(2) hypothèse de variance homogène ?

**Conditions d'application:**  
**Homoscédasticité**  
**Variable normale**



Non ?



Oui ?

Si non, essayer de transformer les variables

# Test de régression linéaire

## Etape 2: tracer la droite et la fonction de régression linéaire

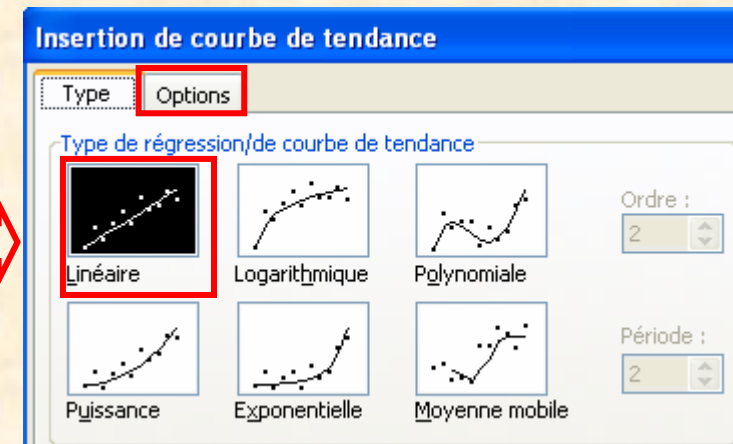
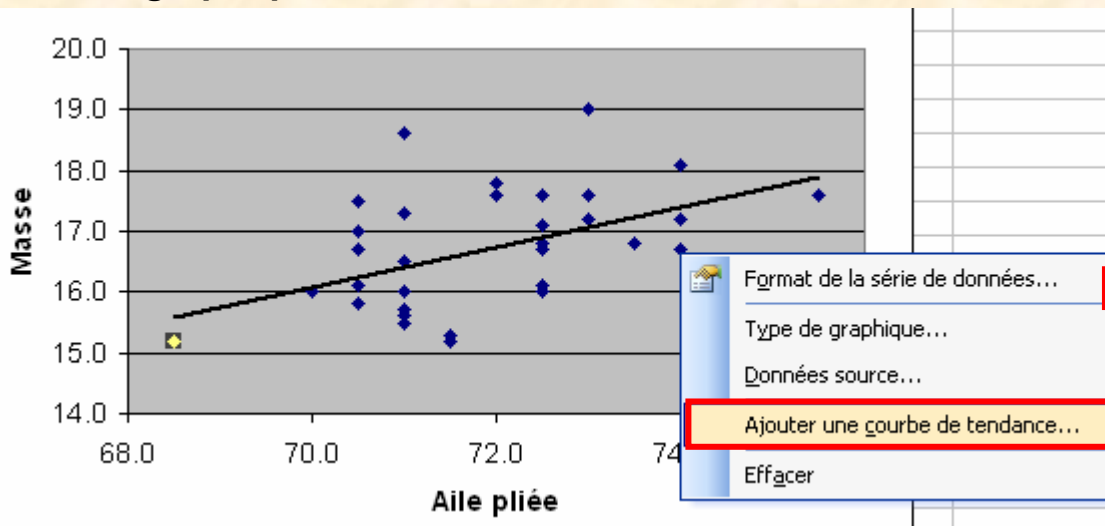
Cliquer sur points  
dans le graphique

Clic droit

Ajouter courbe

Choisir "Linéaire"

Cliquer sur "Options"





# Test de régression linéaire

Etape 2: tracer la droite et la fonction de régression linéaire

**Insertion de courbe de tendance**

Type Options

Nom de la courbe de tendance

☒ Automatique : Linéaire (MA)

☐ Personnalisé :

Prévision

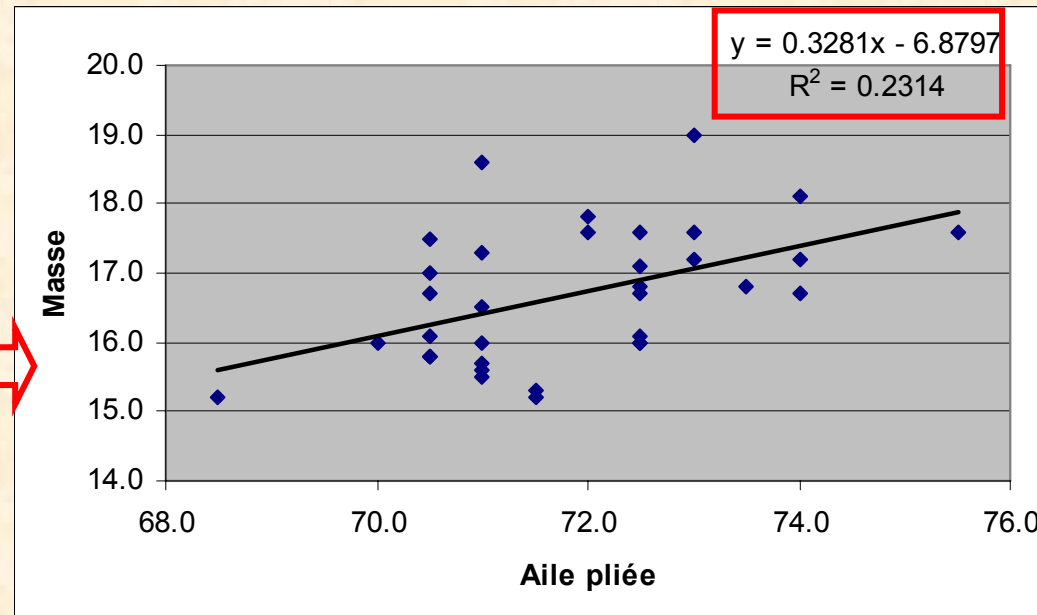
Prospective : 0 unité(s)

Rétrospective : 0 unité(s)

☐ Coupe l'axe horizontal (X) à : 0

☒ Afficher l'équation sur le graphique

☒ Afficher le coefficient de détermination ( $R^2$ ) sur le graphique



Modèle statistique linéaire  
 $\text{Masse} = 0.328 \times \text{LP} - 6.880$

$R^2$  = coefficient de détermination =  
pourcentage de variation expliquée

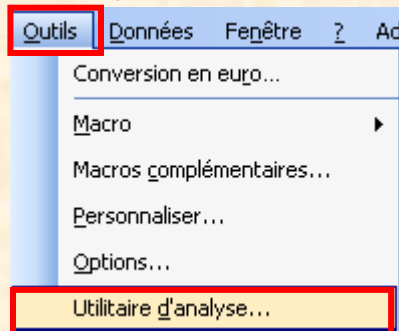
# Test de régression linéaire

Etape 3: test des paramètres de la droite de régression linéaire  
Modèle statistique linéaire

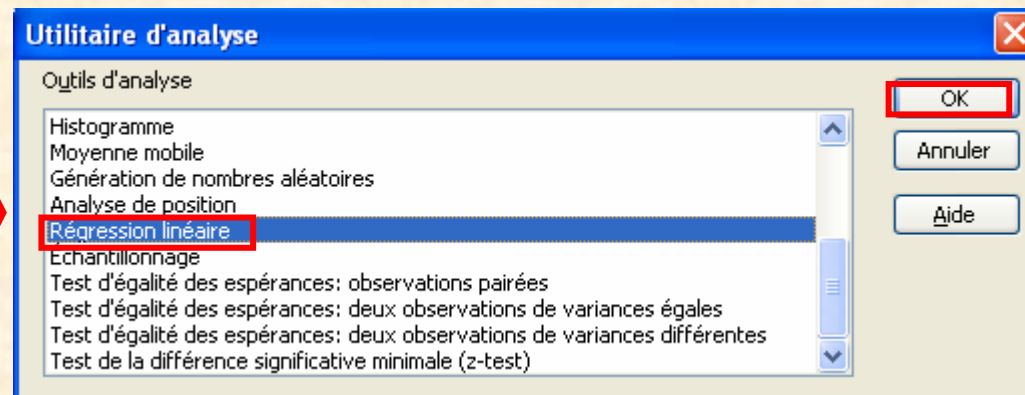
$$\text{Masse} = 0.328 \times \text{LP} - 6.880$$

Validité de cette formule ?  
a et b  $\neq 0$  ?

Ouvrir l'utilitaire  
d'analyse



Choisir "Régression linéaire"



# Test de régression linéaire

## Etape 3: test des paramètres de la droite de régression linéaire

### Définir l'analyse à effectuer

Y=MA	X=LP
15.2	68.5
16.0	70.0
17.5	70.5
16.7	70.5
17.0	70.5
15.8	70.5
15.8	70.5
16.1	70.5
17.3	71.0
18.6	71.0
16.0	71.0
15.7	71.0
15.5	71.0
15.6	71.0
16.5	71.0
15.3	71.5
15.2	71.5

**Régression linéaire**

**Paramètres d'entrée**

Plage pour la variable Y:

Plage pour les variables X:

☒ Intitulé présent ☐ Intersection à l'origine

☐ Niveau de confiance  %

**Options de sortie**

☐ Plage de sortie:

☒ Insérer une nouvelle feuille:

☐ Créer un nouveau classeur

**Analyse des résidus**

☒ Résidus ☒ Courbes des résidus

☐ Résidus normalisés ☐ Courbes de régression

**Probabilité normale**

☐ Diagramme de répartition des probabilités

OK Annuler Aide



# Test de régression linéaire

Etape 4: interprétation des coefficients de la relation

<i>Statistiques de la régression</i>	
<b>Coefficient de détermination multip</b>	<b>0.481 (= coefficient de corrélation)</b>
<b>Coefficient de détermination <math>R^2</math></b>	<b>0.231 (= % de variation expliquée)</b>
Coefficient de détermination $R^2$	0.207
Erreur-type	0.870
Observations	34

**La régression linéaire corrélant la masse à la longueur d'aile explique 23.1% des variations de masse observées**

# Test de régression linéaire

Etape 5: test de la significativité de la relation linéaire

## ANALYSE DE VARIANCE

	Degré de liberté	Somme des carrés	Variance des carrés	F	Valeur critique de F
Régression	1	7.296	7.296	9.634	0.004
Résidus	32	24.234	0.757		
Total	33	31.530			

La relation linéaire entre la masse et la longueur d'aile est significative

$$(F_{1,32} = 9.634, P = 0.004)$$

# Test de régression linéaire

Etape 6: à titre indicatif, significativité et précision des paramètres de la régression linéaire

$$Y = a \times X + b \rightarrow \text{Masse} = a \times \text{LP} + b$$

TEST DE LA SIGNIFICATIVITE (valeur non nulle) DES PARAMETRES DE LA REGRESSION LINEAIRE

		Coefficients	Erreur-type	Statistique t	Probabilité	Seuil de confiance pour seuil de confiance = 95%	Seuil de confiance pour seuil de confiance = 95%
Constante	= b	-6.880	7.597	-0.906	0.372	-22.355	8.596
X=LP	= a	0.328	0.106	3.104	0.004	0.113	0.543



# Test de régression linéaire

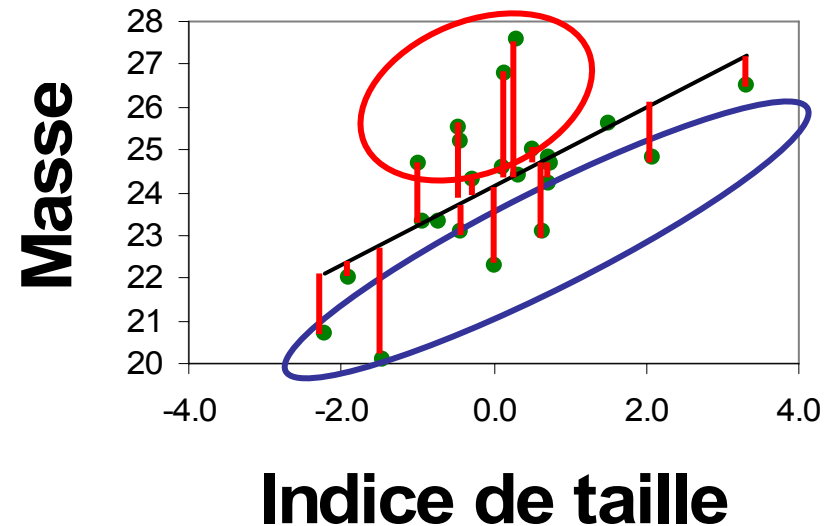
## Bonus 1: utilisation des résidus de la régression linéaire

➤ Rapport entre **masse** et **taille**

=> résidus = **indice de condition corporelle**

A comparer entre:

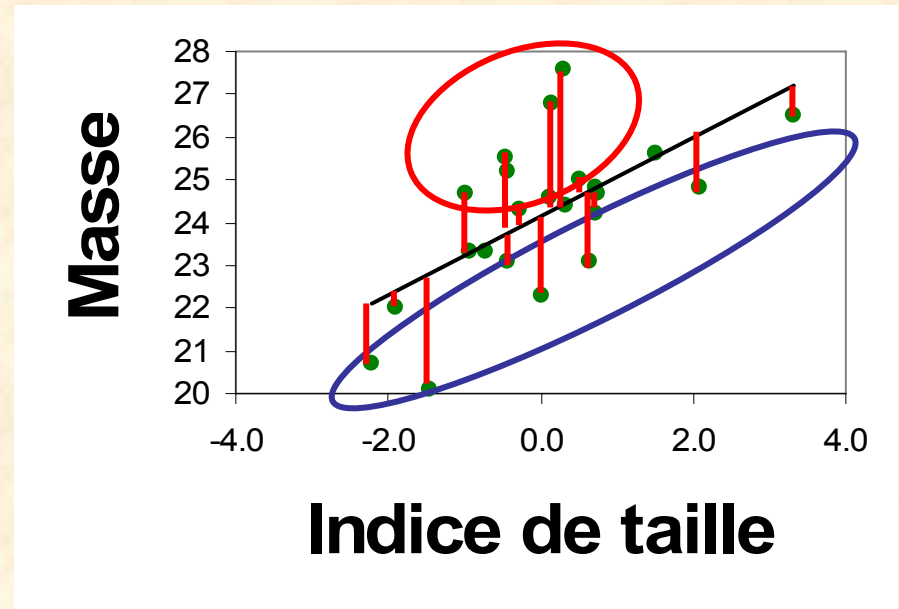
- sites,
- espèces,
- période de l'année



# Test de régression linéaire

## Bonus 1: utilisation des résidus de la régression linéaire

Observation	révisions Y=M	Résidus
1	15.597	-0.397
2	16.089	-0.089
3	16.253	1.247
4	16.253	0.447
5	16.253	0.747
6	16.253	-0.453
7	16.253	-0.453
8	16.253	-0.153
9	16.417	0.883
10	16.417	2.183
11	16.417	-0.417
12	16.417	-0.717
13	16.417	-0.917
14	16.417	-0.817
15	16.417	0.083
16	16.581	-1.281
17	16.581	-1.381
18	16.745	0.855
19	16.745	1.055
20	16.909	-0.209
21	16.909	0.191



# Test de régression linéaire

## Bonus 2: exploration du respect des conditions d'application de l'analyse avec les résidus

**Condition d'application:**

**Homoscédasticité & Normalité**

-> distribution homogène de part et d'autre de 0 des résidus tout au long de l'axe X

**Dans ce cas d'hétéroscédasticité**

-> transformation Log:  $Y' = \ln(Y)$

