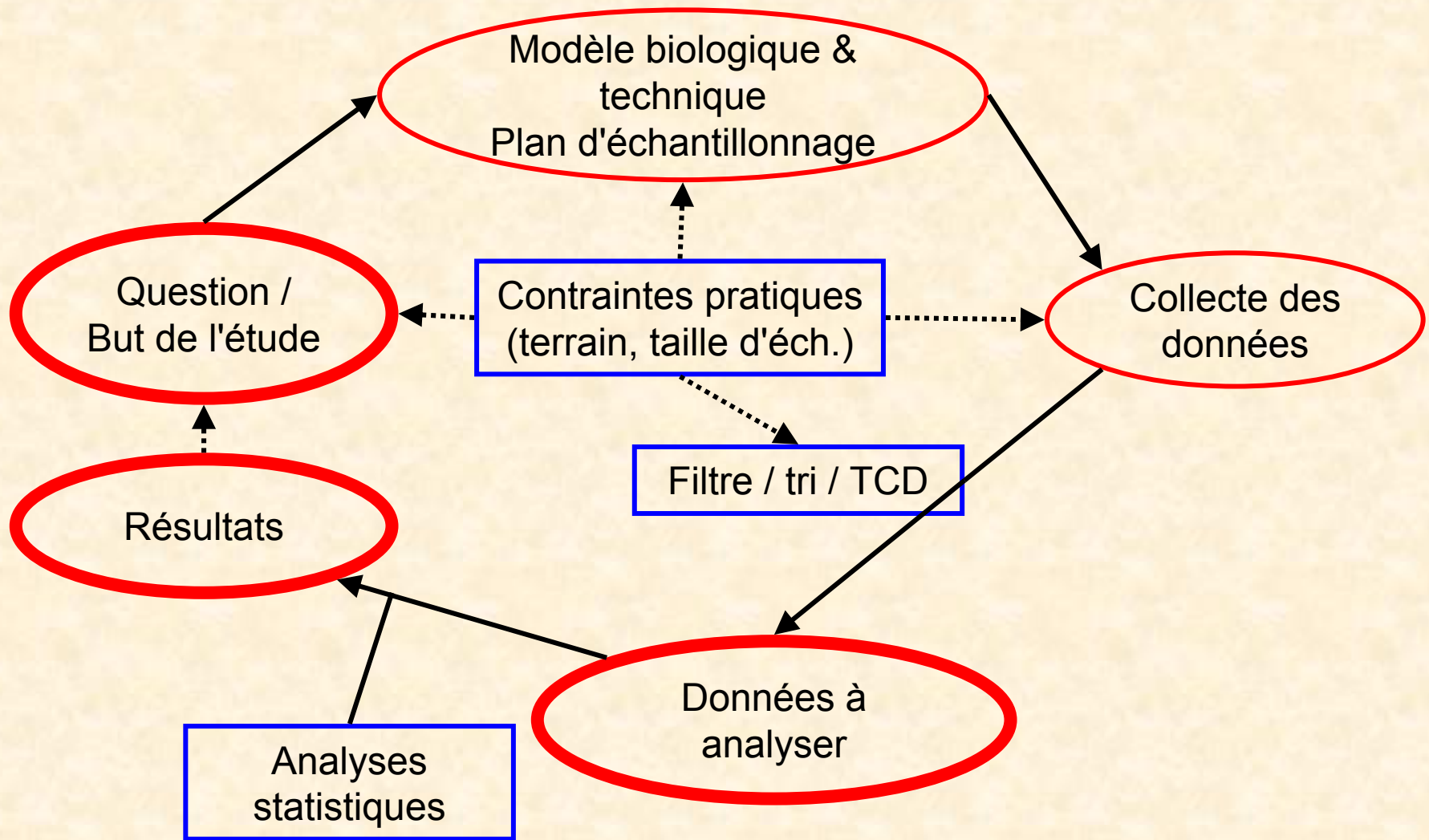


Déroulement d'une analyse et concepts statistiques associés

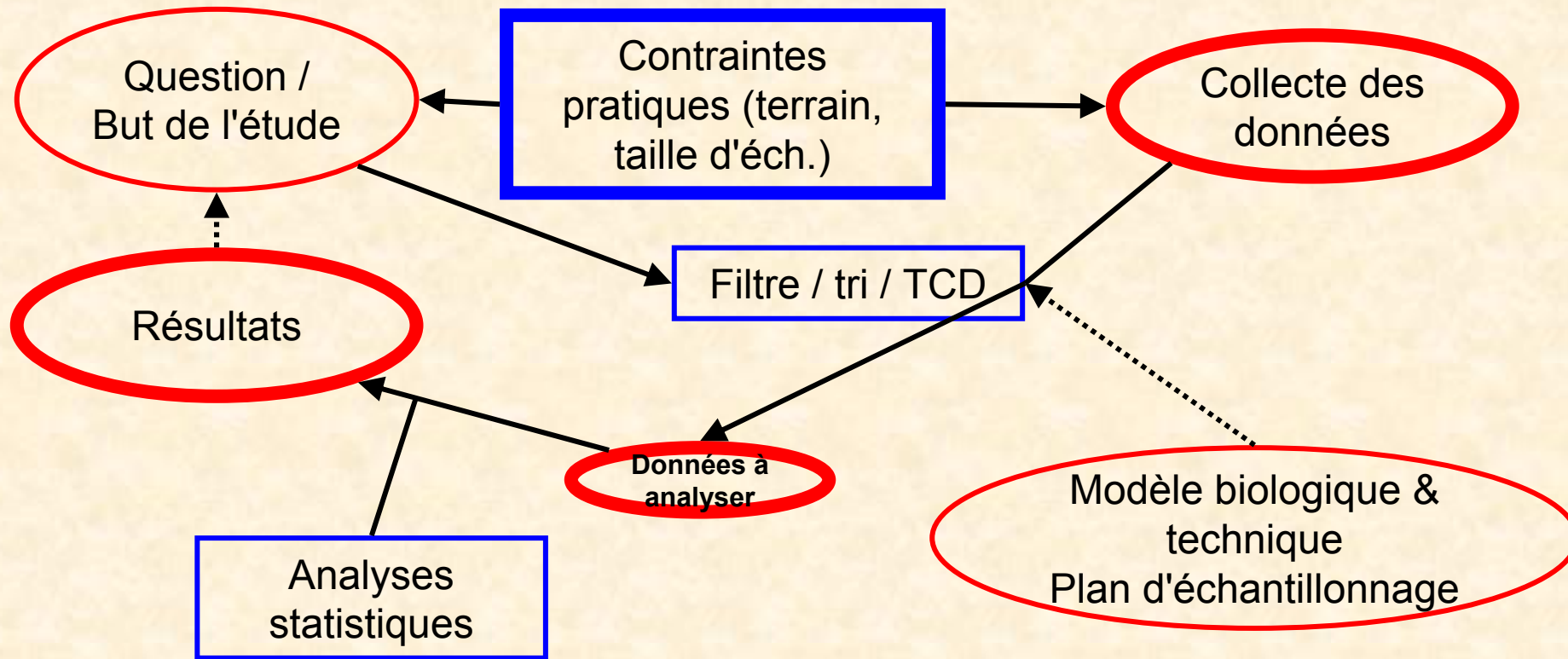
Pierre-Yves Henry

henry@mnhn.fr

Scénario idéal



Scénario plus courant



Principales étapes d'une analyse

(1) Buts de l'analyse ?

- quelles questions biologiques ?
- les effets de quels facteurs sur ces paramètres ?

(2) Quelles données ?

- quels paramètres biologiques ?
- les effets de quels facteurs sur ces paramètres ?
- plan d'échantillonnage ?
- comment collecter les données ?
- quel tri des données (cf. filtre, TCD) ?

(3) Quelles analyses ?

- principe d'un test statistique
- choix du test statistique
- Mise en pratique des tests

(4) *Etre content de soi et s'ouvrir une bière !!! Ou être encore plus content, et retourner sur l'ordi pour affiner l'analyse*



Buts de l'étude



Buts de l'analyse



1- Buts de l'étude: Pourquoi ?

- Différence de **"qualité"** entre sites ?
- Impact d'une **action de gestion** (changement temporel ou spatial) ?
- **Fonctionnement** d'une population ?
- ***Etc.***



1- Buts de l'étude: Pourquoi ?

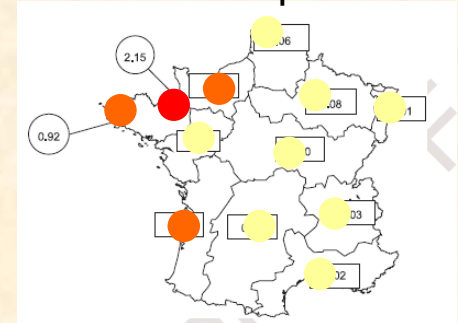
- Comparaison d'**indicateurs de qualité**
 - **âge-ratio** <-> taux de **renouvellement** des individus
 - **indices de condition corporelle** (masse/taille, adip.)
 - indice de **fécondité** (nb jeunes / nb adultes)
 - **taux de retour** (auto-contrôles inter-annuels)
 - taux de **parasitisme**

1- Buts de l'étude: Pourquoi ?

➤ Paramètres de **fonctionnement** des populations

- **sex-ratio** -> migration différentielle entre sexes
- stratégies de **mue**
- **parasitisme** (tiques sur turdidés)
- voies de **migration** (ACROLA)
- *etc.*

% ACROLA par site



Taux de survie, recrutement, croissance de population, recapture, & tailles de population -> stage niveau 2 (CMR)

Quelles données ?

- quels paramètres biologiques ?
- quels effets ?
- plan d'échantillonnage ?
- (comment collecter ?)
- quel tri des données ?

2- Quelles données: quels paramètres ?

Présence, âge, sexe -> **Comptages par catégorie**



Taille, masse, mue, nb. tiques -> **Mesures**



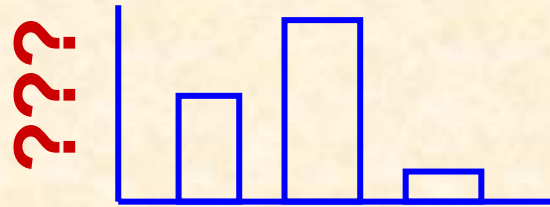
2- Quelles données: quels paramètres ?

Quelles "variables à expliquer" ?

Quelles "variables réponses" ?

Quelles "variables dépendantes" ?

Quels "Y" ?



(1) Variables "catégorielles" (ou discrètes)

A quelle catégorie appartient cet individu ?

- prob. d'être mâle: N mâles *versus* N femelles
- prob. d'être recapturé sur site : N recapturés un an après *versus* N pas recapturés

(2) Variables continues (ou mesurées)

Quelles est la grandeur de cet individu ?

- Taille, Masse
- Nb de tiques

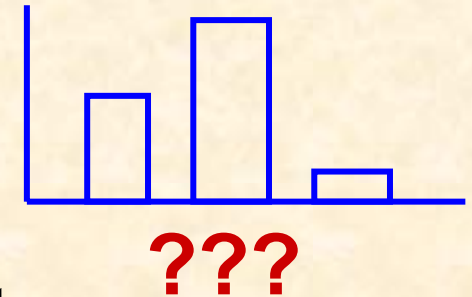
2- Quelles données: quels effets ?

Quelles "variables explicatives" ?

Quelles "variables indépendantes" ?

Quels facteurs ? Quels effets ?

Quels "X" ?



(1) Variables catégorielles

- Sexe, Age
- variations spatiales: Site, Habitat
- variations temporelles: Mois, Année, Température
- variations d'état: Etat parasitaire (O/N)

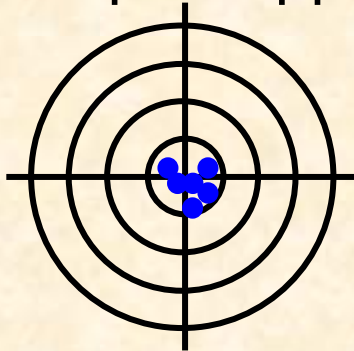
(2) Variables continues

- Masse, taille, coloration
- Nb de tiques

2- Quelles données: plan d'éch. ?

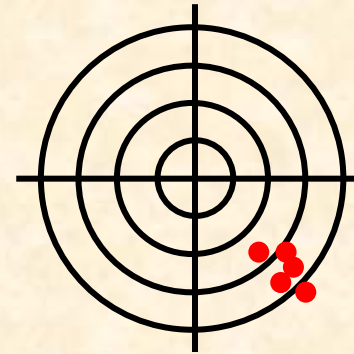
Plan d'échantillonnage = *combien* de données dans le *temps* et dans l'*espace* ?

But 1: que les données soient "représentatives" de la zone, la période, l'espèce étudiée = **que les données ne soient pas biaisées** par rapport au but de l'étude



**Non biaisé
(représentatif)**

Nichoirs sur toute la zone d'étude
Sans repasse
Sans mangeoire
Avec peson calibré
Toute colonie



**Biaisé
(non représentatif)**

tous au bord des chemins
Avec repasse
Avec mangeoire
Avec peson non calibré
Bord de colonie

Exhaustif: OK (mais très coûteux en effort)



Aléatoire = OK



Systématique: OK

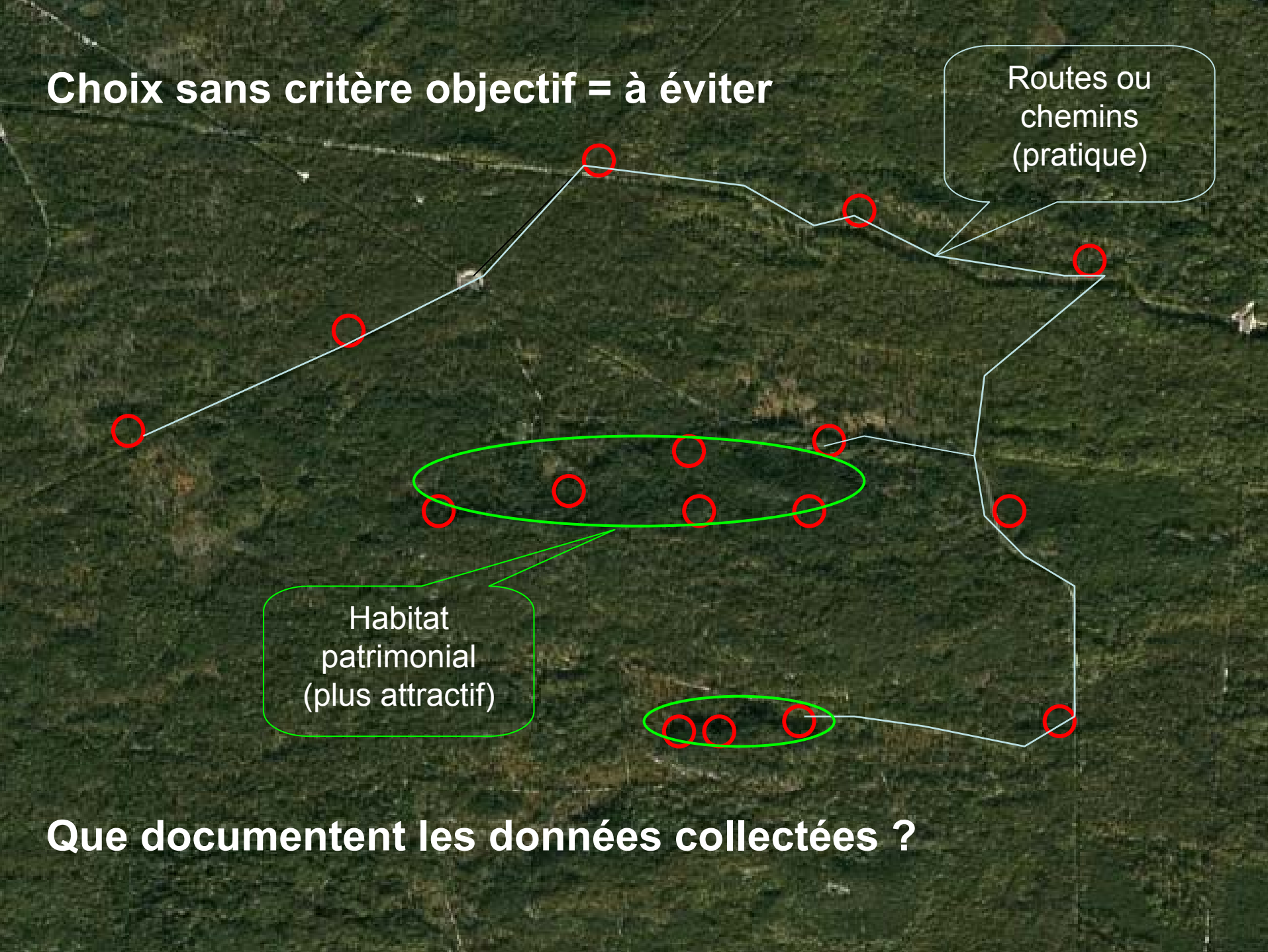


Choix sans critère objectif = à éviter

Routes ou
chemins
(pratique)

Habitat
patrimonial
(plus attractif)

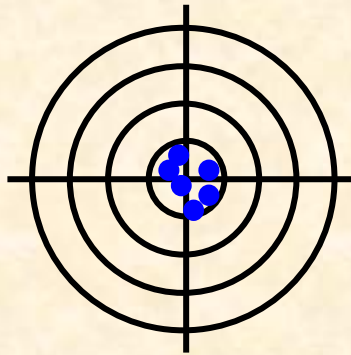
Que documentent les données collectées ?



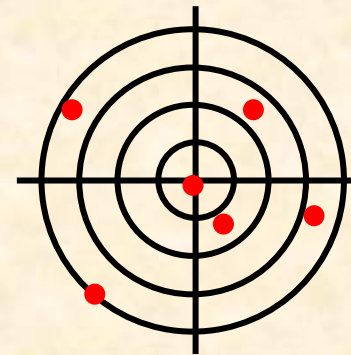
2- Quelles données: plan d'éch. ?

Plan d'échantillonnage = *combien* de données dans le *temps* et dans l'*espace* ?

But 1: que les données apportent des réponses **précises**



Précis



Peu précis

Précision est fonction:

Se mesure par l'**écart-type**
(ou l'**erreur standard** ou la **variance**)

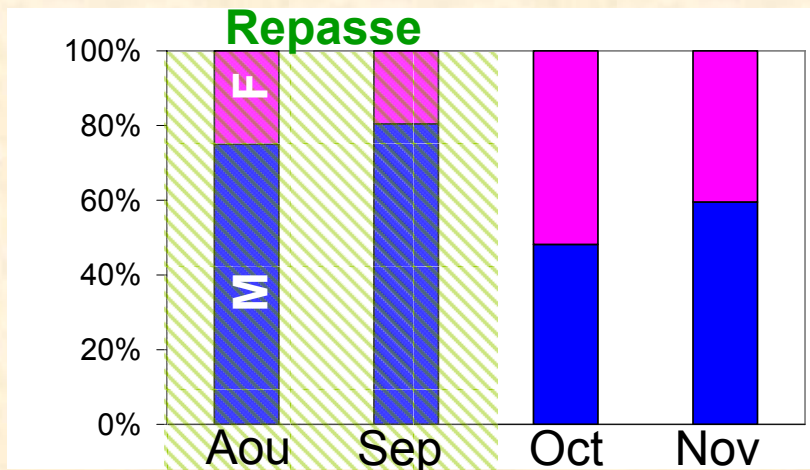
- de la **précision de mesure** (p. ex. taille, balance)
- **variabilité naturelle** (p. ex. masse, aile pliée)
- **nombre d'échantillons**

2- Quelles données: filtrage des données?

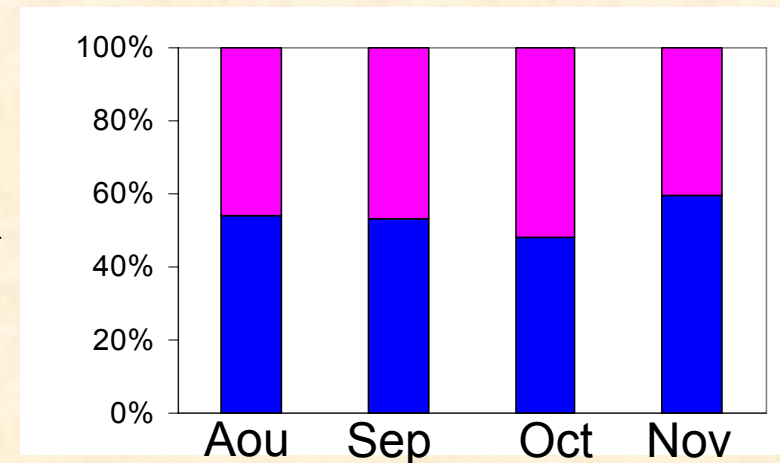
But du filtrage des données:

- > **analyser par morceau** (p. ex. par espèce)
- > pour **compenser des défauts du plan d'échantillonnage**
- > produire une **table contenant les données** nécessaires pour quantifier les paramètres biologiques et les effets retenus

Ex. But de l'étude = test de la variation saisonnière de la sex-ratio



Filtre:
RE = 0



Sex-ratio **varie** entre mois ($\chi^2_3 = 16.5$, $P < 0.05$)

Sex-ratio **constante** ($\chi^2_3 = 1.3$, $P > 0.05$)

Principe d'un test statistique



3- Principe d'un test statistique

Tester si les hypothèses proposées lors de la définition des buts de l'étude sont soutenues statistiquement par les données

Question: La stratégie migratoire de l'Accenteur alpin est-elle différente entre sexes ?

A tester: la sex-ratio hivernale est fonction de l'altitude



-> H_0 = sex-ratio **ne** diffère **pas** avec altitude

-> H_1 = la sex-ratio diffère avec l'altitude

Sachant les données collectées, quelle est la probabilité que H_0 soit vraie ?



3- Principe d'un test statistique

Test de l'adéquation entre une "hypothèse" et les données
Principe d'un test:

**Calculer la probabilité (P) que l'hypothèse
par défaut (dite nulle, H_0) soit vraie**

Probabilité que "l'aile des mâles = aile des femelles"

Probabilité que "le taux de retour dans zone brûlée = zone contrôle"

Probabilité que "la sex-ratio soit la même tout au long de la saison de migration"

**Si l'hypothèse *par défaut* est improbable,
alors l'hypothèse opposée (H_1) est probable**

Aile des mâles \neq aile des femelles

Taux de retour dans zone brûlée \neq zone contrôle

La sex-ratio varie au cours de la saison de migration



3- Principe d'un test statistique

Quels éléments composent un test statistique ?

- (1) Conditions d'application du test
- (2) Statistique
- (3) Nombre de degrés de liberté (ddl, df)
- (4) Probabilité



3- Principe d'un test statistique

Quels éléments composent un test statistique ?

(1) Conditions d'application du test

Spécifiques à chaque test

Evaluer si elles sont à peu près vérifiées avant de prendre en compte le résultat du test

(2) **Statistique** = une variable mesurant l'écart entre la valeur attendue sous l'hypothèse H_0 et la valeur calculée avec les données

(3) **Nombre de degrés de liberté** (ddl, df) = le nombre de comparaisons indépendantes effectuées pour le test

Quand nb d'échantillons \uparrow \rightarrow la statistique \uparrow

Nb de ddl permet de prendre en compte cet **effet du nombre de comparaisons** sur la valeur de la statistique



3- Principe d'un test statistique

Quels éléments composent un test statistique ?

(4) Probabilité que la valeur obtenue pour la statistique s'écarte de 0 simplement par hasard

Quelle est la **probabilité (P)** d'observer cette valeur de la **statistique** avec n **ddl** si l'hypothèse nulle est vraie ?

Quelle est la **probabilité (P)** de se tromper en disant que "la différence est significative"?

Arbitrairement, si **$P < 0.05$**
H0 est considérée **improbable**
⇒ la **différence** est **significative**

Parfois noté * $\leftrightarrow P < 0.05$
 ** $\leftrightarrow P < 0.01$
 *** $\leftrightarrow P < 0.001$

Implique, en moyenne, que
dans 5 tests sur 100 ont
conclue que
H0 est **fausse alors qu'elle**
est vraie !!!

3- Principe d'un test statistique

Quels éléments composent un test statistique ?

(4) Probabilité que la valeur obtenue pour la statistique s'écarte de 0 simplement par hasard

Quelle est la **probabilité (P)** d'observer cette valeur de la **statistique** avec **n ddl** si l'hypothèse nulle est vraie ?

Quelle est la **probabilité (P)** de se tromper en disant que "la différence est significative"?

Arbitrairement, si **$P < 0.05$**

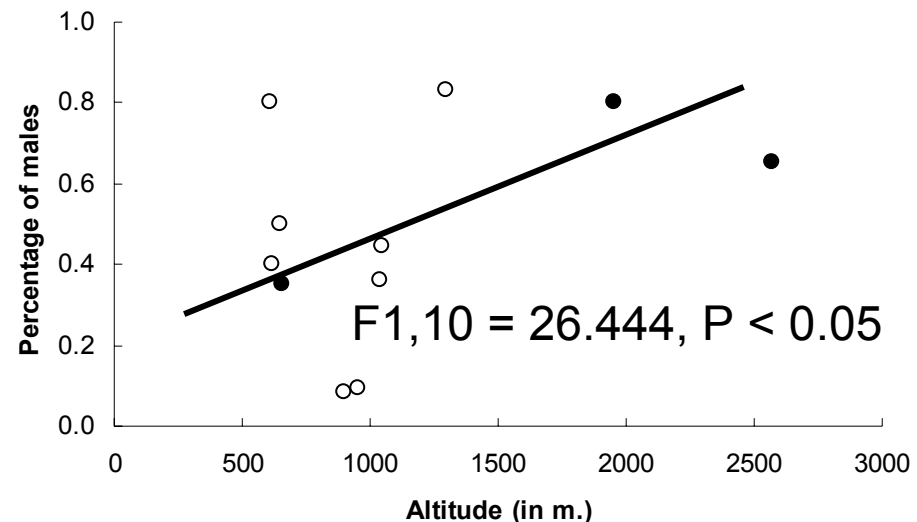
H_0 est considérée **improbable**

⇒ la **différence est significative**

Parfois noté * $\leftrightarrow P < 0.05$

** $\leftrightarrow P < 0.01$

*** $\leftrightarrow P < 0.001$





3- Deux règles fondamentales

1- Une différence non-significative n'implique pas qu'il n'y a pas de différence

2- Corrélation n'est pas causalité

Quelle analyse statistique ?





4- Quelle analyse ?

Fonction de la nature des paramètres et effets

(1) Variables "catégorielles" (ou discrètes)

A quelle catégorie appartient cet individu ?

(2) Variables continues (ou mesurées)

Quelles est la grandeur de cet individu ?



4- Quelle analyse ?

Y \ X	Var. catégorielle	Var. continue
Var. catégorielle		
Var. continue		



4- Quelle analyse ?

Y \ X	Var. catégorielle	Var. continue
	Var. catégorielle	Var. continue
Var. catégorielle	Analyse de fréquence	Faire des catégories Transformer en variable quantitative
Var. continue	ANOVA Comparaison de moyenne	Corrélation Régression

4- Quelle analyse ?

<div>Y</div> <div>X</div>	Var. catégorielle	Var. continue
Var. catégorielle	Age-ratio par année Taux de retour par site Taux de tiques par espèce	Classes de couleur par taille
Var. continue	Aile pliée par sexe Masse par site Durée de séjour par site Date d'arrivée par âge	Masse fonction de taille Nb de tiques et masse Masse fonction de t°