

SECTION 8

Analyse des données



Analyse des données

- **Ne pas retarder l'analyse une fois que les données sont disponibles.**
- **Un programme de statistique ne remplace pas un statisticien.**
- **Ne pas communiquer les résultats susceptibles de changer.**

Il est tentant de penser qu'ayant planifié, organisé et effectué toutes les activités sur le terrain, et ayant rempli une base de données, l'enquête est terminée et l'équipe de prospection peut se relaxer. C'est faux. Il est important de continuer à répondre aux questions posées au début de l'enquête en réalisant une analyse approfondie des données. Il est dangereusement trompeur de penser que de simples tableaux récapitulatifs suffiront à cette fin. Compte tenu de l'investissement entrepris dans l'enquête, une analyse faible et sous-informée, qui produit des résultats douteux, comporte de la fausse économie. Les quatre étapes ci-après devront être réalisées:

Étape 1. Revoir le programme pour l'analyse

Étape 2. Analyser les données

Étape 3. Ajouter de la valeur en se connectant aux bases de données complémentaires

Étape 4. Rédiger le rapport de l'analyse

Ces directives ne décrivent pas la méthodologie statistique dans le détail, car d'autres publications sur ce sujet sont disponibles. En outre, il est important que l'analyse des données soit supervisée par un statisticien qualifié et expérimenté, qui adapte l'analyse aux objectifs de l'enquête plutôt que de suivre une série établie d'instructions.

ÉTAPE 1. REVOIR LE PROGRAMME POUR L'ANALYSE

Le programme pour l'analyse des données devrait avoir été envisagé au cours de la phase de planification (section 4), mais il devrait être revu immédiatement avant le début de l'analyse et confirmé ou modifié selon les besoins.

Choix du calendrier de l'analyse par rapport à la collecte des données

L'analyse devrait être réalisée dès que possible après l'achèvement des activités sur le terrain et de la saisie des données. Il existe plusieurs raisons.

- L'élan créé par les activités précédentes n'a pas encore diminué, et les parties prenantes seront concentrées sur les questions de l'enquête et attendront avec impatience les résultats.
- Il est utile de conduire l'analyse lorsque la planification et les activités de terrain sont encore présentes dans l'esprit de ceux qui y étaient engagés. La documentation des phases précédentes du travail ne sera peut-être pas toujours aussi claire qu'elle le devrait.



- Au sein de l'équipe statistique, la continuité entre la planification et l'analyse est importante, car la conception statistique devrait naturellement conduire aux modèles et aux approches utilisés pour l'analyse.
- Plus l'analyse sera réalisée rapidement, plus les résultats seront communiqués rapidement et plus les actions conséquentes seront mises en œuvre rapidement, plus le rapport coûts-avantages de l'enquête augmentera.

Ressources affectées à l'analyse

Cette section est probablement la plus technique des directives, par conséquent il sera peut-être difficile pour les lecteurs n'ayant pas de compétences en statistiques d'associer le sujet à leur propre expérience. Cet aspect, ajouté au fait que l'expérience dans la conduite des enquêtes sur les ressources zoogénétiques est limitée, risque de sous-estimer sérieusement le temps nécessaire pour l'analyse des données. Le temps nécessaire dépendra de la complexité de la base de sondage, du nombre de questions que l'enquête cherche à affronter, et du nombre des variables nécessaires pour obtenir les réponses à ces questions. Compte tenu que l'analyse sera réalisée en utilisant des ordinateurs, le temps nécessaire pour compléter le travail ne dépendra pas beaucoup, en principe, de la taille de l'échantillon. Il est cependant possible, s'il est nécessaire de vérifier la qualité des données rétroactivement, qu'il existe un élément du travail affecté par la taille de l'échantillon.

Il est probable que l'équipe d'analystes sera de plus en plus efficace avec le temps car elle acquiert de la familiarité avec les approches statistiques utilisées pour répondre aux questions de l'enquête, des notions plus approfondies sur les questions associées à la fiabilité et à l'interprétation des données et une plus grande connaissance des moyens pour accéder aux sources des données auxiliaires qui pourraient ajouter de la valeur. Par conséquent, s'il faut engager une équipe d'analystes dans une série d'enquêtes, l'analyse des données de la première enquête prendra peut-être plus de temps que celles des autres. Toutefois, les dernières enquêtes demanderont peut-être du travail additionnel pour expliquer les implications des changements intervenus dans la conception de l'enquête ou les événements significatifs qui ont affecté la production animale dans la zone à l'étude actuelle par rapport à la précédente. Le temps nécessaire pour aborder ces questions équilibrera certains des avantages de l'expérience.

ÉTAPE 2. ANALYSER LES DONNÉES

Ajustement du modèle

Les objectifs auront piloté la conception (base de sondage) de l'enquête et la base de sondage, à son tour, dirigera les modèles et les approches statistiques qui sont requis pour l'analyse.

Données manquantes

Comme il a été énoncé à la section 7, les procédures de vérification des données révéleront peut-être des erreurs qui sont impossibles à corriger. Deux approches à ce problème peuvent être suivies pendant l'analyse des données:

- les entrées erronées peuvent être traitées comme des données manquantes; ou



- sur la base des connaissances acquises des données de l'enquête, les valeurs moyennes prévues pour les entrées en question peuvent être estimées et saisies – une procédure appelée imputation (pour de plus amples détails, voir FAO, 1995).

D'un point de vue statistique, la première approche est à préférer, à condition que des méthodes statistiques soient utilisées pour surmonter le problème des données manquantes de façon flexible et en minimisant la perte des données, c'est-à-dire sans effacer en bloc d'autres aspects des données. L'utilisation de ces méthodes impliquera peut-être souvent l'imputation de valeurs, mais cette imputation est réalisée de façon optimale. Lorsque la seconde approche est utilisée, la base de données doit enregistrer les points d'information qui sont imputés et ceux qui sont réels, et les méthodes statistiques doivent prendre correctement en compte le fait que les données rassemblées n'étaient pas aussi complètes qu'elles le semblaient. Par exemple, à moins que la question ne soit adressée correctement, les intervalles de confiance dérivés seront plus courts que convenu. Lors de la rédaction du rapport d'analyse, il est important de décrire le traitement utilisé pour les valeurs manquantes.

Programmes statistiques

Les algorithmes informatiques nécessaires pour ajuster les modèles statistiques sont disponibles dans une vaste gamme de programmes génériques de statistique, et certains programmes visent de façon spécifique les données numériques de l'enquête. Il est conseillé d'utiliser des programmes très estimés car leurs algorithmes sont plus largement compris et acceptés, et il est plus probable qu'ils ne contiennent pas d'erreurs de programmation. Le programme choisi doit être en mesure de gérer la taille de la base de données qui sera produite dans l'enquête. Ce détail devra être vérifié avec les fournisseurs de logiciels. Il faudrait prévoir les futures enquêtes et s'assurer que le programme s'adapte à des modèles qui visent tant les enquêtes de suivi que de base. En cas de doutes, il faudrait rechercher les conseils de professionnels spécialisés dans ce secteur.

Le fait qu'un programme puisse traiter les données et produire des «résultats» ne remplace pas les connaissances des professionnels. Les modèles disponibles dans le programme ne seront peut-être pas adéquats ou ne produiront pas les estimations efficaces des valeurs recherchées dans l'enquête – d'autres modèles assureront peut-être des estimations de meilleure qualité avec des intervalles de confiance plus courts. L'existence d'un programme de statistique qui peut s'adapter à des modèles spécifiques *influencera peut-être* les choix de conception au stade de la planification, mais il ne faudrait pas sacrifier la conception pour réaliser l'analyse avec un programme particulier.

Le mouvement des logiciels analytiques «libres» est en hausse, ce qui représentera peut-être une alternative aux programmes commerciaux. Un exemple est «R» (voir par exemple, Lumley, 2004). L'utilisation de «R» a l'avantage de permettre le partage des outils entre des groupes différents, tout en les personnalisant. Il s'agit peut-être d'une thématique de discussion entre les équipes responsables des différentes enquêtes.

Transmettre les «résultats émergents» aux partenaires

Pendant l'analyse, il sera peut-être souhaitable de transmettre les «résultats émergents» aux promoteurs ou partenaires importants de l'enquête, même si certaines incertitudes



ENCADRÉ 46

Le rapport sur l'analyse des données – liste de contrôle

1. Les objectifs de l'enquête sont-ils énumérés?
2. La base de sondage et d'autres méthodologies de terrain ont-elles été décrites de façon adéquate?
3. La description des méthodes statistiques est-elle adéquate, notamment les méthodes utilisées pour effacer ou transformer les données et pouvoir les répéter sans l'aide d'un statisticien?
4. Le choix des modèles est-il justifié?
5. Les hypothèses étayant les modèles sont-elles énumérées? (Les références à d'autres publications peuvent être suffisantes.)
6. Les modèles:
 - s'adressent-ils aux objectifs; et
 - reconnaissent-ils toutes les variables de la conception?
7. Les résultats décrits dans le rapport:
 - visent-ils les objectifs;
 - fournissent-ils des mesures fiables pour les estimations et les prédictions réalisées;
 - évaluent-ils les biais possibles et leur ampleur potentielle;
 - évaluent-ils les effets que des hypothèses de modèles invalidants pourraient avoir sur les résultats;
 - identifient-ils les sources importantes d'erreurs expérimentales et, le cas échéant, comment ces erreurs seront-elles peut-être mieux contrôlées pendant de futures enquêtes;
 - signalent-ils les résultats inexplicables qui justifieraient peut-être d'autres recherches pour décider s'ils sont le produit d'une faille de la conception ou des résultats inattendus mais «réels»; et
 - identifient-ils les actions d'urgence qui seront peut-être nécessaires en raison des résultats de l'analyse?
8. A-t-on envisagé comment personnaliser la présentation des résultats pour les différentes parties prenantes?

demeurent sur les modèles qui ont produit les résultats. Le besoin de transmettre ces résultats aura peut-être plusieurs raisons, notamment la nécessité de:

- conserver l'intérêt des parties prenantes;
- fournir des preuves d'activité aux promoteurs;
- inciter des actions d'urgence; ou
- obtenir d'autres informations ou données de suivi pour adresser un résultat inattendu (potentiellement dissiper le doute et décider si le résultat est le produit d'une faille de la conception ou un résultat «réel»).



Dans ces circonstances, les rapports contenant les résultats «émergents» devraient toujours comprendre des avertissements déclarant qu'ils sont préliminaires et sujets à changement. Les rapports devraient être toujours classés comme confidentiels, et traités ainsi jusqu'à ce que l'équipe de prospection, dans son ensemble, et le groupe de travail chargé d'élaborer la stratégie les ont analysés. La publication prématurée de résultats, qui sont ensuite rétractés, transmettra des messages équivoques aux parties prenantes, ce qui créera de la confusion et un manque de confiance dans les résultats définitifs de l'enquête. Cet aspect compromettra tout l'investissement en temps, en efforts et en fonds consacrés à l'enquête et représentera une menace pour les enquêtes futures.

ÉTAPE 3. AJOUTER DE LA VALEUR EN SE CONNECTANT AUX BASES DE DONNÉES COMPLÉMENTAIRES

D'autres données, rassemblées à de fins différentes, seront peut-être utiles pour interpréter les résultats de l'enquête et pour ajouter de la valeur à ses produits. Il sera peut-être intéressant d'utiliser les informations biogéographiques pour associer la gamme ou la densité de races particulières aux caractéristiques du climat, du terrain (altitude, type de sol, caractéristiques de la surface, etc.) ou de la flore, ou à l'épidémiologie.

ÉTAPE 4. RÉDIGER LE RAPPORT DE L'ANALYSE

Le rapport sur l'analyse devrait être détaillé, mais compréhensible pour les membres non spécialistes de l'équipe de prospection et du groupe de travail chargé d'élaborer la stratégie. La responsabilité de communiquer les résultats au-delà de ces groupes est abordée à la section 9. Il est recommandé, pour rendre compréhensible le rapport, d'énoncer les détails techniques aux annexes. Ceci devrait permettre à un lecteur informé, mais non pas expert, d'apprécier la qualité technique de l'analyse et de comprendre l'importance des résultats pour la gestion des ressources zoogénétiques sans devoir analyser tout le matériel technique. L'encadré 46 présente une liste de contrôle pour adapter le rapport sur l'analyse. Ce rapport devra faire partie du rapport final de l'enquête (voir section 9). Il est recommandé de rédiger également les résultats de l'enquête dans un document à soumettre à une revue scientifique, par exemple le journal *Ressources génétiques animales*¹⁷.

Lors de la rédaction du rapport sur l'analyse, il sera utile de commencer à réfléchir à des systèmes pour résumer les résultats détaillés à présenter aux différents groupes de parties prenantes, en tenant compte de leurs différentes préoccupations et capacités. Par exemple, quelles statistiques de résumé seront nécessaires? Quelles figures graphiques exprimeront au mieux les résultats? Les stratégies de communication sont décrites à la section 9, mais il faudrait réfléchir rapidement aux utilisations possibles des résultats pour économiser ensuite beaucoup de temps et d'efforts.

¹⁷ <http://journals.cambridge.org/action/displayJournal?jid=AGR>



