

Article

Symposium 2008 :
Collecte des données : défis, réalisations et nouvelles orientations

Enquête auprès des entreprises complexe menée uniquement sur le Web – Expériences et enseignements tirés

par Johan Erikson

2009



Enquête auprès des entreprises complexe menée uniquement sur le Web – Expériences et enseignements tirés

Johan Erikson¹

Résumé

Ces dernières années, d'importants progrès ont été réalisés dans le domaine de la collecte des données en ligne. Aujourd'hui, un grand nombre de bureaux de la statistique offrent une option de réponse en ligne pour de nombreux types d'enquête. L'opinion selon laquelle la collecte de données en ligne peut accroître la qualité des données tout en réduisant les coûts de collecte est très répandue. L'expérience a montré que les entreprises auxquelles a été offerte l'option de remplir un questionnaire en ligne au lieu d'un questionnaire imprimé ont mis du temps à adopter l'option en ligne. Par ailleurs, les expériences ont également montré qu'en favorisant la réponse en ligne plutôt que l'utilisation d'un questionnaire imprimé, il est possible d'accroître le taux d'adoption de l'option en ligne. Cependant, les études décrivant ce qui se passe lorsque la stratégie de prise de contact est modifiée radicalement et que l'option de réponse en ligne est la seule offerte dans une enquête auprès des entreprises complexe sont encore rares. En 2008, Statistics Sweden a décidé d'utiliser une stratégie axée à peu près uniquement sur le Web dans l'enquête sur la production industrielle (PRODCOM). Le questionnaire en ligne a été élaboré au moyen de l'outil généralisé de conception d'enquêtes en ligne utilisé par l'organisme. Le présent article décrit la solution en ligne et certaines expériences relatives à l'enquête PRODCOM de 2008, y compris les données de traitement sur les taux de réponse et les ratios d'erreurs, ainsi que les résultats d'un suivi cognitif à l'enquête. Certaines importantes leçons apprises sont également présentées.

Mots clés : Collecte de données en ligne, stratégies de collecte, taux d'adoption, taux de réponse.

1. Contexte

Au cours de la dernière décennie, de nombreux bureaux de la statistique ont fait de grands efforts en vue d'adopter la collecte des données en ligne. Les raisons qui rendent ce mode de collecte attrayant sont nombreuses. Étant donné les fortes pressions politiques en vue de réduire le fardeau de réponse imposé avant tout et par-dessus tout aux entreprises, l'offre d'options de réponse électroniques a été considérée comme un moyen facile d'arriver à réduire le fardeau de réponse tant réel que perçu. Deuxièmement, les bureaux de la statistique espèrent qu'en intégrant des fonctions telles que les contrôles de vérification et les fonctions d'aide, ils amélioreront la qualité des données recueillies en ligne et réduiront peut-être le coût de l'exécution de la collecte des données, puisque aucune saisie de données ne serait nécessaire, que les coûts d'envoi par la poste pourraient être réduits, etc.

En réalité, les répondants ont mis du temps à adopter les options de réponse en ligne. Par conséquent, ces dernières années, de nombreux efforts ont été faits en vue d'accroître le ratio de répondants choisissant l'option Web. Dans certaines enquêtes, il a été décidé d'adopter la démarche audacieuse de n'offrir à peu près que l'option de réponse en ligne. Toutefois, si les taux d'adoption sont faibles, les avantages hypothétiques pourraient s'avérer limités ou inexistantes. Par ailleurs, d'aucuns s'inquiètent des effets que pourrait avoir sur les taux de réponse globaux la promotion de l'option en ligne uniquement et le fait de « cacher » l'option papier. Par conséquent, il est intéressant d'étudier les effets de la démarche consistant à adopter à peu près la réponse « en ligne uniquement ».

¹Johan Erikson, Statistics Sweden, PCA/LEDN, SE-701 89, Sweden, johan.erikson@scb.se

2. Expérience antérieure

2.1 Stratégies de collecte des données à Statistics Sweden

Statistics Sweden collecte des données en ligne depuis plus de cinq ans. Avant cela, des questionnaires électroniques étaient offerts dans certaines enquêtes sous forme de tableurs Excel ou d'autres formes semblables. Au début des années 2000, quelques programmes d'enquête ont intégré des outils personnalisés pour la collecte des données sur Internet. En 2004, des travaux ont débuté en vue de créer un outil généralisé de collecte des données en ligne, dont la première version a commencé à être utilisée en avril 2005. Depuis, les fonctionnalités et l'utilisation de cet outil se sont beaucoup étendues. Même s'il a été créé au départ pour les enquêtes auprès des entreprises, son usage a été étendu à tous les types d'enquêtes, c'est-à-dire auprès des entreprises, des administrations publiques, des ménages et des particuliers. Aujourd'hui, plus de 130 enquêtes différentes ont utilisé l'outil généralisé.

Auparavant, la plupart des enquêtes suivaient ce que l'on peut décrire comme une stratégie d'« offensive papier » pour prendre contact avec les répondants auxquels étaient envoyés une lettre de couverture et un questionnaire imprimé. L'option leur était donnée de répondre en ligne s'ils le souhaitaient, en utilisant un numéro d'identification d'utilisateur et un mot de passe normalement imprimés sur le questionnaire papier (les lettres de couverture étant génériques et non personnalisées, il est plus facile d'imprimer l'information propre aux répondants sur les questionnaires). Les trousse de rappel envoyées contenaient un nouveau questionnaire imprimé. Quand cette stratégie est utilisée, de 5 % à 20 % des répondants choisissent ordinairement l'option en ligne, le taux d'adoption étant plus élevé pour les entreprises que pour les particuliers.

Les deux dernières années, les stratégies ont commencé à évoluer et à s'orienter davantage vers une « offensive en ligne ». Il existe plusieurs versions d'offensives en ligne, mais les plus courantes consistent à ne pas envoyer de questionnaire imprimé dans la trousse de rappel ou, en allant une étape plus loin, à inclure uniquement une lettre de couverture contenant le numéro d'identification d'utilisateur et le mot de passe dans l'envoi initial. Dans le cas des enquêtes auprès des administrations publiques, la stratégie d'offensive en ligne est celle qui est maintenant la plus fréquemment utilisée, tandis que pour les enquêtes auprès des particuliers et des ménages, la stratégie classique fondée sur un questionnaire papier demeure la norme. Dans le cas des enquêtes auprès des entreprises, la stratégie varie d'une enquête à l'autre, avec une tendance à évoluer de plus en plus vers la stratégie fondée sur la réponse en ligne.

2.2 Diriger les répondants vers l'option Web

Statistics Sweden a effectué assez peu d'études sur les effets de l'adoption de stratégies axées davantage sur la réponse en ligne. Holmberg, Lorenc et Werner (2007) ont étudié les incidences dans le cas d'une enquête sur la santé menée auprès des particuliers. En ce qui concerne les enquêtes auprès des entreprises, une expérience a été effectuée en 2007 (données pour 2006) dans le contexte d'une enquête longitudinale sur les traitements et salaires (« LINDA »). Dans l'expérience, une stratégie d'offensive en ligne a été utilisée pour un groupe d'entreprises, tandis qu'une stratégie classique a été utilisée pour les autres. L'expérience et les premiers résultats ont été présentés dans Erikson (2007), ainsi que les résultats obtenus pour quelques autres enquêtes. Les résultats définitifs de l'expérience peuvent être consultés (en suédois) dans Erikson (2008). Les figures 2.2-1 et 2.2-2 illustrent les principaux résultats de l'expérience.

Figure 2.2-1
Taux de réponse au cours du temps dans l'expérience LINDA de 2006

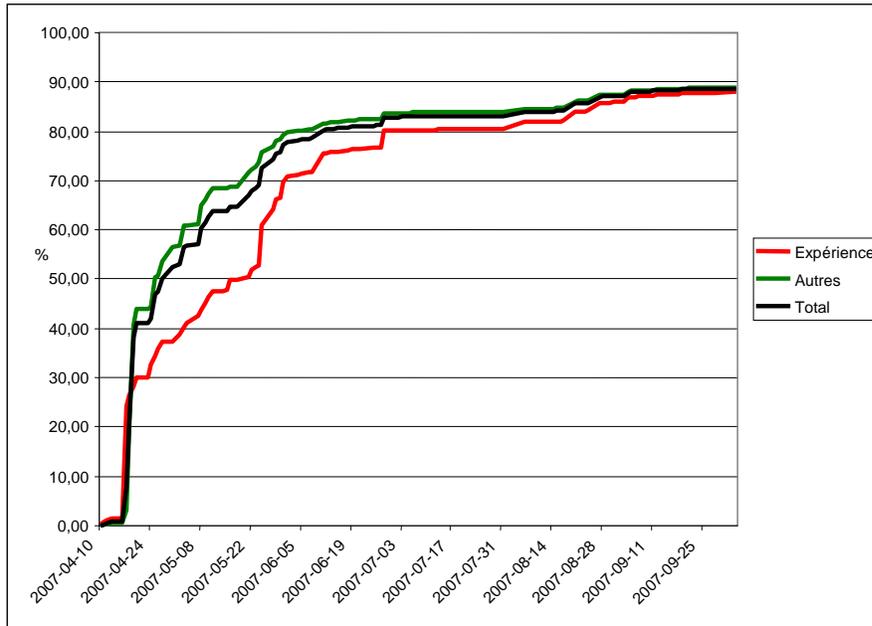
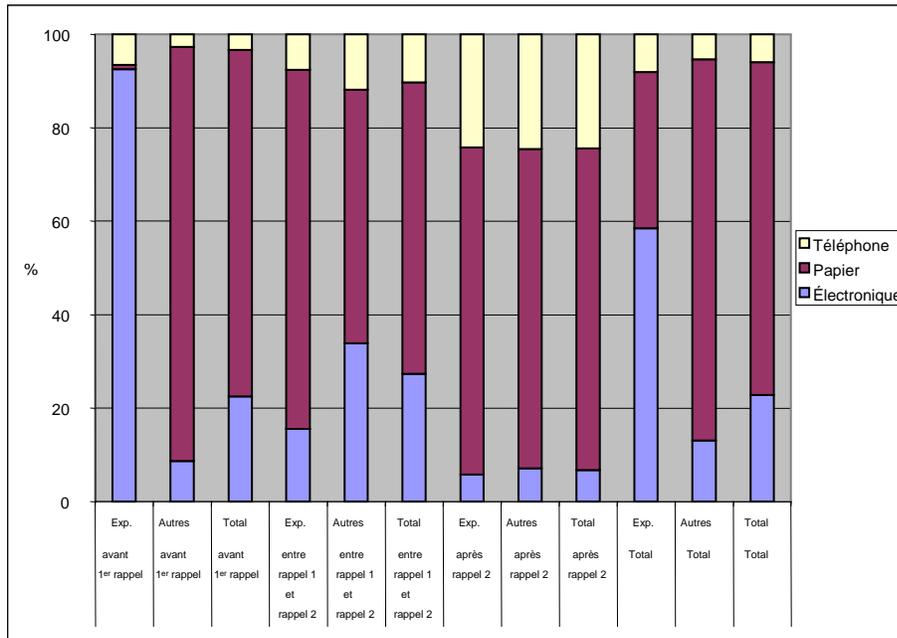


Figure 2.2-2
Taux d'adoption de la réponse en ligne dans l'expérience LINDA de 2007



La figure 2.2-1 montre qu'en fin de compte, l'expérience n'a eu qu'un effet très faible sur le taux de réponse total. Ce qu'il a été intéressant d'observer est l'« effet d'attente » qui s'est manifesté dans le groupe expérimental. Il semble que certaines entreprises ont attendu le rappel, qui contenait un questionnaire imprimé, pour répondre. Cela s'observe également dans la figure 2.2-2, où le taux d'adoption de la réponse en ligne est nettement plus élevé dans le groupe expérimental avant le rappel. La figure 2.2-2 indique aussi clairement que le mode choisi par les répondants dépend de la stratégie, le taux d'adoption d'Internet étant beaucoup plus élevé pour le groupe expérimental, ainsi que des périodes durant lesquelles l'option en ligne est promue (avant le 1^{er} rappel pour le

groupe expérimental, et après le 1^{er} rappel pour les autres groupes, qui n'ont pas reçu un nouveau questionnaire imprimé avec le rappel).

3. Enquête menée « uniquement sur le Web »

En se basant sur l'expérience LINDA, il serait intéressant d'examiner une enquête pour laquelle il a été décidé d'offrir à peu près la réponse en ligne seulement, en utilisant la stratégie d'offensive en ligne, afin de voir si les résultats de l'expérience seront reproduits. Mais il serait tout aussi intéressant de voir si les coûts de la collecte des données sont réduits et si la qualité des données est meilleure qu'auparavant. En 2008, cette démarche a été faite pour l'enquête sur la production industrielle.

3.1 Caractéristiques de l'enquête

L'enquête sur la production industrielle est une enquête européenne menée par tous les pays membres de l'Union européenne. Elle porte sur la production de biens et de services industriels répartis selon le type de biens ou services. Elle est régie par le règlement PRODCOM (production communautaire) et, par souci de simplicité, nous l'appellerons enquête PRODCOM dans la suite de l'exposé. En Suède, cette enquête est annuelle, tandis qu'elle est réalisée à plus court terme dans certains autres pays européens. Voici quelques caractéristiques intéressantes de l'enquête :

* Il s'agit d'une enquête à participation obligatoire, ce qui est le cas de nombreuses enquêtes réglementées. En théorie, les répondants ne peuvent tout simplement pas refuser de fournir les données, même si le mode offert n'est pas celui qu'ils préfèrent.

* La production doit être déclarée selon le type de biens ou de services, conformément à une nomenclature de plus de 10 000 biens et services, de sorte que la déclaration des données est une tâche très détaillée et très complexe.

* Les données doivent être déclarées en valeurs et en quantités de biens ou services. Pour de nombreuses entreprises et de nombreux biens ou services, le concept de quantité est très difficile à saisir. Beaucoup d'entreprises ne comptabilisent pas les quantités dans leur système de tenue de livres. En plus, pour de nombreux biens et services, il n'existe en réalité aucune quantité pertinente à communiquer. Cela signifie que la collecte de ces données prend beaucoup de temps et que de nombreuses reprises de contact doivent avoir lieu avec les entreprises pour poser des questions ou pour recueillir les données manquantes.

* L'enquête est un recensement au-delà du seuil de 20 employés (10 dans certains secteurs), ce qui signifie que le taux de renouvellement est très faible et que les répondants connaissent l'enquête.

En 2008, la stratégie de collecte de l'enquête PRODCOM était la suivante : l'envoi original par la poste ne contenait qu'une lettre de couverture indiquant le numéro d'identification et le mot de passe à utiliser. Après cinq semaines, un rappel contenant un questionnaire imprimé a été envoyé.

3.2 Taux de réponse et taux d'adoption

Les figures 3.2-1 et 3.2-2 donnent les principaux résultats de l'enquête PRODCOM de 2008. Le taux de réponse total n'a pas été affecté négativement du tout, étant même un peu plus élevé en 2008 qu'en 2007. En outre, le taux d'adoption de la réponse en ligne a été très élevé, 90 % des entreprises qui ont répondu avant le rappel ayant choisi cette option, et jusqu'à 50 % l'ayant fait même après avoir reçu le questionnaire imprimé. En tout, le taux d'adoption de la réponse en ligne était d'environ 70 %. Et, ce qui est peut-être l'élément le plus intéressant, aucun effet d'attente n'a été observé, le profil de réponse au cours de la période de collecte étant presque identique à celui de l'année précédente. L'une des explications possibles pourrait être la conception de l'enquête, dans laquelle les taux de renouvellement sont faibles et les répondants sont expérimentés. Cela pourrait signifier qu'ils savent qu'ils doivent répondre et qu'ils le font à un moment commode, qui est à peu près le même chaque année. Le taux élevé d'adoption de la réponse en ligne montre qu'il est possible d'utiliser la collecte de données en ligne même dans le cas de questionnaires longs et complexes, parce qu'autrement, le taux d'adoption après l'envoi du questionnaire imprimé devrait être nettement plus faible.

Figure 3.2-1
Taux de réponse au cours du temps dans l'enquête PRODCOM de 2008

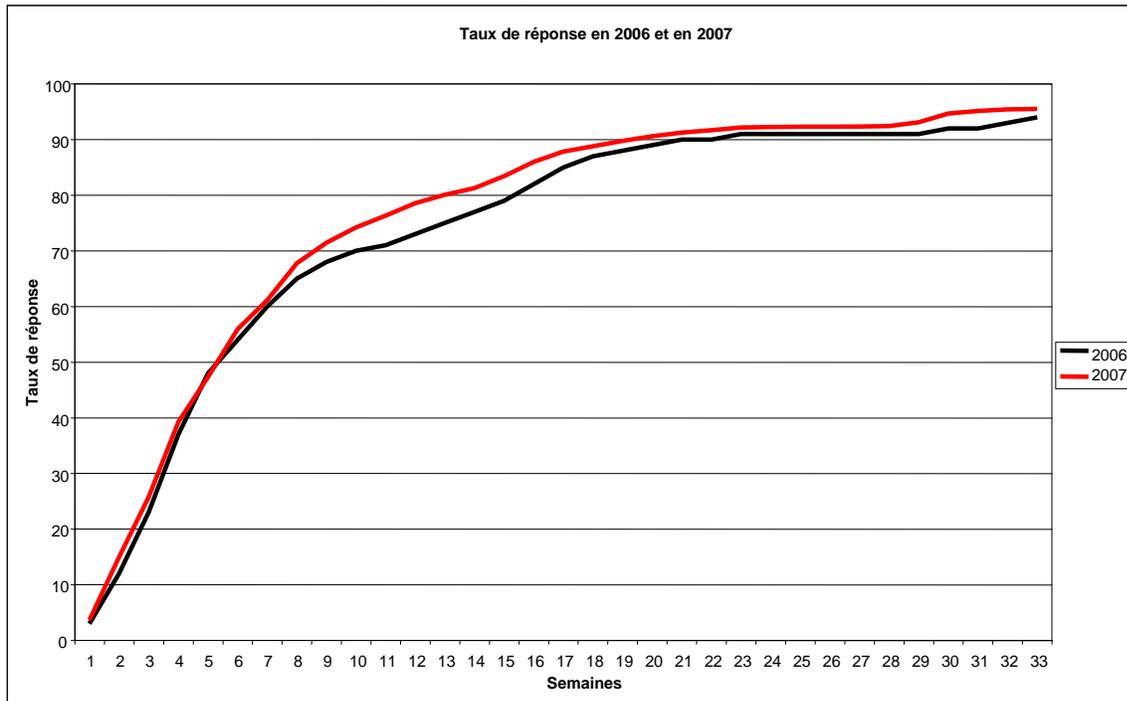
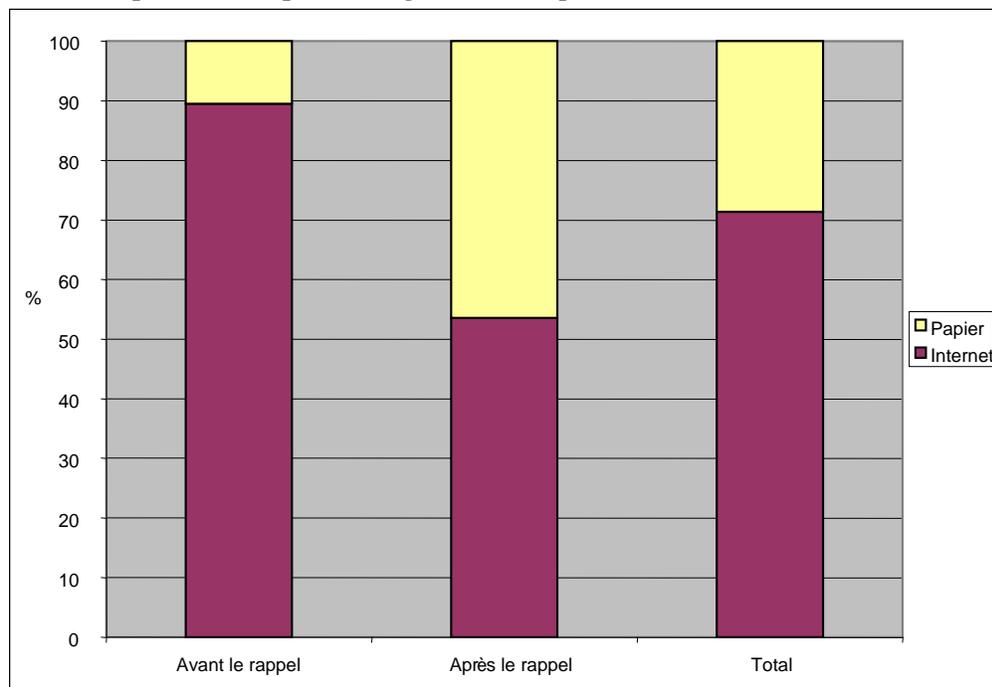


Figure 3.2-2
Taux d'adoption de la réponse en ligne dans l'enquête PRODCOM de 2008



3.3 Coût de la collecte des données et qualité des données

Alors que les taux de réponse et les taux d'adoption étaient bons, les résultats en ce qui concerne le coût et la qualité des données étaient tout au plus satisfaisants. Le coût de l'envoi par la poste a été réduit d'un tiers parce que le questionnaire initial n'a pas été envoyé par la poste, mais les coûts de main-d'œuvre pour les processus de collecte et de vérification des données sont demeurés plus ou moins les mêmes que l'année précédente. Le tableau 3.3-1 montre certaines données de traitement pour les trois dernières années de production.

Tableau 3.3-1

Données de traitement pour la collecte des données de l'enquête PRODCOM, 2006 à 2008

	2006	2007	2008
	nombre		
Courriels entrants	278	352	493
Courriels sortants	415	408	1 022
Appels téléphoniques entrants	482	452	548
Appels téléphoniques sortants :			
Pas de réponse	1 088	1 379	1 786
Entente en vue d'un nouvel appel	342	977	631
Prise de contact	1 243	2 126	2 026
Somme des appels téléphoniques sortants	2 673	4 482	4 443
Nombre d'unités contactées	1 346	2 085	2 203
Nombre d'unités dans l'enquête	3 814	4 128	4 818
Taux d'unités contactées (%)	35	51	46

Selon les données de traitement, les nombres d'appels téléphoniques et de courriels entrants étaient à peine plus élevés en 2008 qu'en 2007, ce qui indique que les répondants se sont adaptés assez facilement au questionnaire en ligne. Si la demande de questionnaires imprimés avait été forte ou qu'un grand nombre de répondants avaient éprouvé des difficultés techniques, le nombre d'appels ou de courriels entrants aurait été plus élevé. Cette conclusion est également appuyée par les taux de réponse et d'adoption ne démontrant aucun effet d'attente, ce qui est encourageant. Par ailleurs, si les données entrantes avaient été de meilleure qualité qu'auparavant, le nombre d'appels ou de courriels sortants auraient été plus faible qu'auparavant. Or, même s'il se dégage une légère tendance dans ce sens (contact a été pris avec 46 % du nombre total d'unités comparativement à 51 % l'année précédente), cela n'a pas suffi à réduire le coût total de la phase de collecte. Celle-ci a encore pris autant de temps qu'avant l'introduction du questionnaire en ligne.

Les raisons pour lesquelles les coûts n'ont pas été plus faibles et les données, de meilleure qualité que les années précédentes, sont multiples :

* La raison principale tient à la nature de l'enquête et aux difficultés concernant les quantités. Comme nous l'avons mentionné précédemment, de nombreux répondants éprouvent des problèmes à déclarer les quantités, ce qui signifie que de nombreuses prises de contact avec les entreprises doivent avoir lieu pour obtenir ces données. Les deux dernières années, des efforts ont également été faits en vue d'obtenir des données sur les quantités auprès d'entreprises qui n'avaient pas fourni ce genre de données antérieurement. Il a donc fallu multiplier les prises de contact, ce qui est évident si l'on compare les données de 2006 et de 2007.

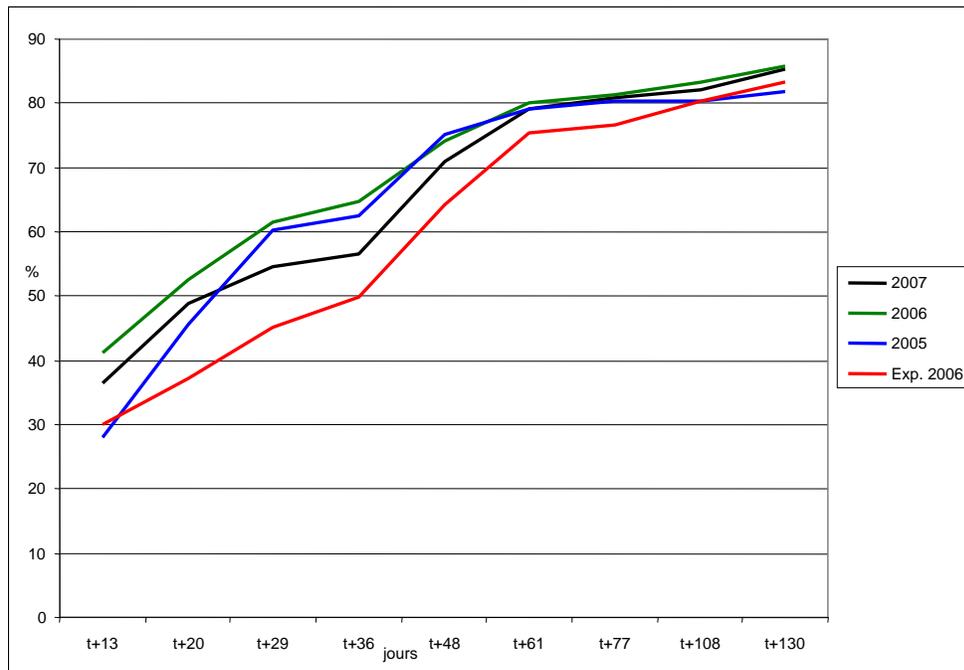
* Lorsque des données détaillées sont demandées dans le cas d'une nomenclature compliquée, le processus de collecte des données est affecté par les changements apportés à cette nomenclature. Le nombre de ces changements varie d'une année à l'autre. En 2008, comme il a été très élevé, des efforts supplémentaires ont dû être faits pour vérifier les données transmises par de nombreux répondants, ce qui a nécessité des prises de contact supplémentaires et occasionné plus de travail. Il est fort possible que les changements de nomenclature aient été l'un des facteurs importants pour lesquels les coûts de collecte n'ont pas été réduits. Il sera intéressant de voir l'année prochaine si des effets plus importants sont observés alors que les modifications à la nomenclature devraient être moindres.

* Le questionnaire en ligne n'était pas d'assez bonne qualité. Lorsque l'on introduit des questionnaires en ligne, il est fréquent de les concevoir de manière à ce qu'ils ressemblent autant que possible à leur équivalent imprimé, même si l'on adopte une stratégie d'offensive en ligne. En outre, les gestionnaires d'enquête hésitent souvent à ajouter un trop grand nombre de fonctionnalités intégrées, telles que les contrôles de vérification, parce qu'ils ne veulent risquer que les répondants choisissent de ne pas répondre du tout s'ils font face à des rejets à la vérification. Dans le questionnaire PRODCOM, une meilleure disposition du questionnaire et un plus grand nombre de fonctionnalités intégrées, par exemple le rappel aux répondants que des quantités devraient être déclarées, pourrait signifier l'obtention de meilleures données originales. Ne pas utiliser les possibilités d'Internet réduit la probabilité de réduire les coûts et d'accroître la qualité. Un suivi cognitif à l'enquête PRODCOM (Henningsson et Hartwig 2008) montre que les répondants s'attendent à ce qu'un formulaire en ligne leur fournisse des renseignements et que le questionnaire PRODCOM présente des défauts, tant dans sa version papier que dans sa version en ligne.

4. Autre expérience récente

Compte tenu du succès de l'expérience, le programme d'enquête LINDA a décidé d'utiliser la stratégie d'offensive en ligne et de ne pas inclure de questionnaire imprimé dans le premier envoi par la poste. Ce qui présentait un intérêt tout particulier dans cette enquête, après l'expérience PRODCOM, était de vérifier si l'effet d'attente observé dans l'expérience se manifesterait encore ou si l'on obtiendrait le même profil sans effet d'attente que dans l'enquête PRODCOM. Les résultats sont présentés à la figure 4-1. L'effet d'attente observé dans l'expérience est encore présent, quoique plus faible que l'année précédente. L'une des conclusions que l'on pourrait tirer est que les profils d'adoption de la réponse en ligne diffèrent selon l'enquête et qu'une analyse préalable des caractéristiques de cette dernière est importante. L'enquête LINDA s'adresse dans une beaucoup plus grande mesure que l'enquête PRODCOM à des petites entreprises et est caractérisée par un taux de renouvellement un peu plus élevé. Cela pourrait signifier que les répondants ne sont pas aussi préparés à adopter l'option de réponse en ligne que ceux de l'enquête PRODCOM. L'effet d'attente plus faible que l'année précédente pourrait, à son tour, signifier que les répondants s'habituent à ce que les enquêtes soient réalisées en ligne uniquement et, par conséquent, finissent par utiliser Internet. Il serait intéressant de faire des comparaisons plus directes entre divers types d'enquêtes dans l'avenir. L'existence d'effets d'attente pourrait aussi signifier que l'adoption d'un mode de réponse strictement en ligne, en n'offrant jamais d'option papier, pourrait avoir des effets négatifs sur les taux de réponse et le fardeau de réponse perçu. Cet aspect devrait également être étudié plus en détail.

Figure 4-1
Taux de réponse au cours du temps dans l'enquête LINDA, 2005 à 2007



5. Conclusion

L'étude réalisée pour l'enquête PRODCOM représente une étude de cas d'une enquête auprès des entreprises complexe. Certains résultats sont positifs, dont le taux élevé d'adoption de la réponse en ligne sans aucun effet négatif sur le taux de réponse total, ce qui concorde avec les résultats d'études antérieures et avec les attentes. Par ailleurs, d'autres résultats sont négatifs. Le coût total de collecte était aussi élevé qu'auparavant et la qualité des données ne s'est pas améliorée. Selon une expérience antérieure, on pouvait s'attendre à un coût plus faible et une qualité plus élevée. Les principales raisons de ces résultats négatifs sont la complexité de l'enquête proprement dite, les changements de nomenclature et une quantité insuffisante de renseignements dans le questionnaire en ligne. Les deux derniers facteurs devraient être étudiés l'année prochaine, après une amélioration du questionnaire et quand de moins nombreux changements seront apportés à la nomenclature. Une autre expérience récente montre que les résultats diffèrent dans une certaine mesure selon le type d'enquête. Les travaux devraient se poursuivre en vue d'effectuer des comparaisons directes entre divers types d'enquête.

Bibliographie

- Erikson, J. (2007). Effects of offering web questionnaires as an option in enterprise surveys, *Proceedings of the Third International Conference on Establishment Surveys*, American Statistical Association. 1431-1435.
- Erikson, J. (2008). LINDA-experimentet, rapport inédit, Örebro, Suède: Statistics Sweden.
- Henningson, B. et Hartwig P. (2008). Industrins varuproduktion – Rapport från kognitivt test, rapport inédit, Örebro, Suède: Statistics Sweden.
- Holmberg, A., Lorenc, B. et Werner P. (2007). LOTTA P2 – Optimal kontaktstrategi vid blandad insamling papper och webb, rapport inédit, Örebro, Suède: Statistics Sweden.