

Politique générale de révision et de correction d'erreur

Direction de la recherche, des études, de l'évaluation et des statistiques (DREES)

Synthèse

- La DREES peut être amenée à réviser les valeurs de ses statistiques pour qu'elles rendent compte de ce qu'elles mesurent de la manière la plus fiable possible et pour garantir leur pertinence. Elle peut également être amenée à corriger une erreur que peuvent comporter ses statistiques.
- Révision et correction d'erreur ne se confondent pas, et renvoient à des notions et à des démarches différentes.
- Dans les deux cas, la DREES met en œuvre des procédures établies, normalisées et planifiées suivant les standards internationaux.
- Les révisions comme les corrections d'erreurs sont annoncées selon un protocole adapté à la source de révision ou d'erreur, à leur ampleur, à leurs conséquences pour les utilisateurs et à la sensibilité des statistiques publiées.
- Toutes les explications nécessaires à la compréhension de ces modifications accompagnent la diffusion des sources statistiques concernées sur le site www.drees.solidarites-sante.gouv.fr.

1. Distinguer une révision de la correction d'une erreur de production

1.1 Définition d'une révision

Les **révisions** consistent à modifier la valeur d'une statistique déjà diffusée et non à corriger une erreur de production. La modification apportée peut être légère ou substantielle. Il peut s'agir :

- d'une **révision « de routine »**, régulière, intégrée au processus de production, qui vise à améliorer la fiabilité des statistiques produites. En effet, afin de mettre au plus tôt de l'information à la disposition des utilisateurs, des statistiques peuvent être diffusées alors que les données ne sont pas encore disponibles dans leur intégralité. Par la suite, l'intégration de données collectées dans un délai plus long permet d'améliorer la qualité de l'information produite.

- du **résultat d'une refonte du dispositif de production**, exceptionnelle, induisant une rupture de série, si possible corrigée par une rétopolation¹. Les révisions issues de refontes visent à améliorer la pertinence des données produites en introduisant des changements méthodologiques ou conceptuels. La fréquence de ces modifications est toutefois limitée autant que possible, afin d'éviter l'introduction de ruptures de séries, défavorables à la comparabilité chronologique des données.

Une révision de routine peut résulter par exemple :

- de l'intégration d'informations plus complètes ou plus récentes en entrée du dispositif de production ;
- du calage sur des données de cadrage ;
- de la modification de paramètres d'estimation (coefficients de modèles CVS-CJO par exemple), etc.

Les révisions issues d'une refonte peuvent notamment avoir pour origines :

¹ La rétopolation d'une série consiste, en partant d'un nouveau niveau d'une statistique, estimé pour une ou plusieurs années données, à reconstituer l'ensemble de la série.

- des changements méthodologiques (modification de l'année de base ou de référence, de la méthode de calcul, du champ, de l'échantillon, du questionnaire, du dispositif de production, etc.) ;
- l'utilisation de nouvelles sources ou des modifications des sources ;
- des changements de concepts, de définitions, de nomenclatures ;
- de changements de géographie (nouveau découpage géographique) ;
- une nouvelle réglementation.

1.2 Définition d'une erreur de production

Une erreur est un résultat incorrect², inattendu, découvert après publication, résultant d'un défaut des données sources collectées, d'un écart au processus de production statistique prévu ou d'une anomalie dans la diffusion.

Elle peut porter sur les données elles-mêmes (y compris celles reportées dans un commentaire, un tableau, un graphique, un schéma ou une infographie) mais aussi sur le texte (présentation, explication, analyse) ou les métadonnées les accompagnant (toute indication méthodologique : définition, champ, source, hypothèse, méthode).

Tous les supports de diffusion peuvent être concernés par une erreur : publication papier ou électronique (y compris base de données et visualisation des données), webservice, page du site Internet ou article du blog, publication sur les réseaux sociaux, communiqué de presse, de même que les fichiers détail (mis à disposition en interne ou destinés aux chercheurs), les données livrées à Eurostat et les codes sources.

En résumé, une erreur est un phénomène qui se corrige en appliquant le processus de production prévu ou en l'adaptant à la marge.

1.3 Distinguer une erreur de production d'une révision

Comme une erreur, une révision induit la modification de données ou d'informations déjà publiées. Elle s'en distingue par :

- sa planification : une révision est prévue, annoncée à l'avance³, tandis qu'une erreur est inattendue ;
- son origine : une révision est due à la prise en compte d'une version améliorée de données sources ou à une modification méthodologique⁴, tandis qu'une erreur résulte d'un « accident » dans le processus de production ou de diffusion.

L'incertitude statistique (erreur aléatoire - liée au fait de sonder une partie de la population étudiée seulement) inhérente à tout processus de production statistique, ainsi que les biais de mesure (erreur de couverture, erreur due à la non-réponse, etc.) ne relèvent pas de la présente politique.

Pour distinguer en pratique une erreur d'une révision, la DREES s'appuie, comme l'ensemble de la statistique publique, sur **l'arbre de décision présenté en annexe**.

² C'est-à-dire ne répondant pas aux critères de fiabilité prévalant à la diffusion de la donnée.

³ Sauf dans le cas particulier d'une révision non planifiée due à des changements inattendus dans les données sources.

⁴ Une révision « de routine » est due à la disponibilité de données plus complètes ou de meilleure qualité et une révision « méthodologique » est due à des changements de législation, de concepts, de normes, de méthodes, de sources, de classification, d'année de base.

2. Le processus de révision en pratique

La mise en œuvre d'une révision suit un cheminement logique en 5 étapes : décider, organiser, calculer, documenter et informer. Ce cheminement est illustré par le tableau ci-dessous :

Décider	Organiser
<p>Le chef de la DREES peut décider de mettre en œuvre une révision à condition :</p> <ul style="list-style-type: none"> • qu'elle soit conforme aux dispositions réglementaires nationales ou internationales encadrant la production des statistiques ; • de gagner en pertinence, en fiabilité ; • de respecter les principes de la statistique publique⁵ ; • de permettre de maintenir la cohérence avec d'autres productions de la statistique publique. 	<p>Un calendrier de révision est établi, en prévoyant :</p> <ul style="list-style-type: none"> • la diffusion d'une information provisoire précoce de qualité suffisante ; • une fréquence des révisions limitée, quitte à intégrer en une fois plusieurs causes de révisions ; • l'harmonisation des révisions, au sein d'un même domaine statistique ; • si possible la synchronisation des révisions, au sein d'un même domaine, entre pays, avec le changement de réglementation associé, etc. <p>L'opportunité et la faisabilité d'une rétropolation des séries sont évaluées.</p>
<p>Calculer</p> <ul style="list-style-type: none"> • Les statistiques révisées sont calculées. • Si possible, les séries sont rétropolées 	
Documenter	Annoncer
<p>Dans le cas des révisions de routine, le statut des données est signalé :</p> <ul style="list-style-type: none"> • par exemple selon la norme SDMX : P (valeur provisoire), SD (semi-définitive), Non affiché (définitive) le cas échéant ou R (révisée) ; • si elles concernent tout un tableau, la mention des révisions est portée sous le tableau. <p>Dans le cas d'une refonte, une note méthodologique permet aux utilisateurs d'évaluer et d'interpréter les nouvelles valeurs.</p> <p>Les révisions et leurs effets sur les résultats sont si possible analysés.</p> <p>Les rubriques qualité consultables sur le site de la DREES renseignent sur les révisions (calendrier des révisions, raison, nature, etc.).</p>	<p>Le calendrier (révisions de routine) ou la date (refontes) des révisions est annoncé(e) :</p> <ul style="list-style-type: none"> • en général, dès que possible, et au plus tard lors de la diffusion des révisions ; • dans le cas de modifications méthodologiques d'envergure, le mois précédent. <p>Sont diffusées : les statistiques révisées, les éventuelles données rétropolées, la documentation méthodologique et les rubriques qualité sur les révisions (cf. « documenter »).</p> <p>Les données publiées antérieurement sont éventuellement maintenues un certain temps.</p> <p>Au besoin, un plan de communication accompagnant les révisions est établi.</p>

⁵ Les principes du *Code de bonnes pratiques de la statistique européenne* sont appliqués en France à l'ensemble des statistiques produites par le service statistique public, comprenant l'Institut national de la statistique et des études économiques (Insee) et les 16 services statistiques ministériels (SSM). Ils prévoient notamment une communication active vis-à-vis des utilisateurs, aussi bien en termes d'annonce des changements que de transparence des méthodes et d'aide à la compréhension des révisions.

3. Le processus de correction d'erreur en pratique

En principe, la mise en œuvre d'une correction d'erreur suit un cheminement similaire à celui d'une révision (décider, organiser, calculer, documenter et informer). Elle est toutefois plus complexe, notamment parce qu'elle requiert un travail préalable pour identifier le type d'erreur, la méthode de correction, son coût et la communication à mettre en place pour l'accompagner.

3.1 Erreurs mineures et erreurs majeures

Une erreur est considérée comme :

- **majeure** si elle modifie de façon non triviale et potentiellement spéculative l'interprétation des données ou informations diffusées et présente des **conséquences** (estimées) **significatives pour les utilisateurs** ;
- **mineure** dans les autres cas, notamment si la correction de cette erreur, même importante, est évidente pour le lecteur, qu'elle porte sur un élément de forme (typographie, erreur grammaticale, etc.) ou sur un élément de contenu (données, commentaires, etc.).

Les conséquences de l'erreur ne peuvent pas être parfaitement évaluées *ex-ante*. Elles sont approchées par la combinaison :

- du type d'indicateur touché et de sa sensibilité dans l'analyse économique ou sociale ;
- des usages connus des données concernées et des conséquences anticipées de l'erreur qui en découlent ;
- de la visibilité des données concernées (comme proxy de leur degré de diffusion) ;
- de l'ancienneté de l'erreur au moment de sa découverte.

Le type de l'erreur (mineure/majeure) détermine le traitement de l'erreur ainsi que la communication sur l'erreur et sa correction (voir ci-après).

3.2 La correction rapide d'une erreur est la règle

Conformément au code européen de bonnes pratiques, **les erreurs sont corrigées dans les meilleurs délais**, si possible simultanément sur l'ensemble des supports concernés (voir 1.2).

Du point de vue des utilisateurs, les corrections présentent un gain, mais peuvent également présenter un coût (coût d'intégration des données, de reprise de leurs traitements et analyses). Les avantages et inconvénients doivent être mis en balance, de façon à limiter le nombre et la fréquence de rediffusion de données et informations déjà publiées.

Ainsi, deux principes sont retenus :

- **une erreur majeure** fait l'objet d'une **nouvelle diffusion dans les meilleurs délais** ;
- une erreur mineure peut être corrigée sans délai si la balance avantages/inconvénients est bénéfique pour l'utilisateur. Dans le cas contraire, un certain délai peut être envisagé pour intégrer la correction, notamment si celle-ci peut intervenir à l'occasion d'une nouvelle diffusion (dans le cas de données diffusées fréquemment) ou si la charge de production courante nécessite de différer la prise en charge de la correction, de façon à minimiser le risque de retard dans la mise à disposition des estimations suivantes.

Ne sont pas systématiquement corrigées les erreurs mineures induisant un écart non significatif entre la valeur diffusée d'un indicateur et sa valeur correctement estimée.

4. Information des utilisateurs en cas de correction d'erreur

4.1 En amont de la correction

Lorsque leur correction nécessite un délai, les erreurs majeures sont signalées, en amont de leur correction :

- à un endroit bien visible sur les supports concernés et, dans certains cas, à tout autre emplacement susceptible d'attirer l'attention du public ;
- par un message *ad hoc* destiné aux utilisateurs connus des données (notamment aux abonnés dans le cas d'une erreur sur une publication et dans certains cas aux médias), transmis par les canaux habituels de diffusion.

Le signalement de l'erreur comporte alors :

- la description précise de l'erreur ;
- la date à laquelle la correction sera effectuée (ou a minima un délai maximal) ;
- l'origine de l'erreur (lorsqu'elle a pu être identifiée) ;
- son impact potentiel (lorsqu'il a pu être évalué) : ordre de grandeur, conséquence sur le message transmis, etc. ;
- les supports éventuels donnant des informations plus précises sur l'erreur.

Dans le cas où le **risque de confusion** pour l'utilisateur reste non négligeable malgré ces signalements, **les informations incorrectes peuvent être en outre retirées** du site Internet jusqu'à leur correction.

4.2 Au moment de la correction

Lors de la correction d'une erreur :

- les informations erronées sont remplacées par leur version corrigée et sont définitivement supprimées, sauf suivi particulier (indice des prix à la consommation, etc.) ;
- sauf dans le cas d'erreurs de forme, la version corrigée est accompagnée d'une **mention de la correction** ;
- **les utilisateurs connus sont informés** de la correction des erreurs majeures, selon les modalités décrites au 4.1.

La mention de la correction suit les trois principes suivants :

1/ Elle indique :

- la date à laquelle la correction a été effectuée ;
- l'ampleur de la correction ;
- pour les erreurs majeures, l'explication de l'erreur (origine) et de la correction effectuée.

2/ Elle est **placée à l'endroit de l'erreur ou à proximité immédiate** selon les possibilités (bases de données notamment).

Pour les erreurs majeures, elle est en outre portée à un endroit bien visible sur le support (par exemple, avertissement sur la page de présentation pour une publication électronique) et éventuellement à tout autre emplacement susceptible d'attirer l'attention.

3/ Elle est, dans la mesure du possible, pérenne. C'est le cas, en particulier, pour les publications.

Cas particulier des publications papier :

- Les erreurs mineures de forme sont corrigées uniquement dans les nouveaux exemplaires tirés.
- Les autres erreurs font l'objet :
 - dans les nouveaux exemplaires tirés : d'une mention de la correction effectuée à l'emplacement de l'erreur ;
 - dans les exemplaires déjà tirés non encore distribués : d'un erratum. Les publications papier ne font pas l'objet de retirages.

4.3 Après la correction

Toutes les erreurs (hors erreur de forme), ainsi que leur résolution, sont documentées (nature, origine, correction effectuée) et archivées de façon centralisée.

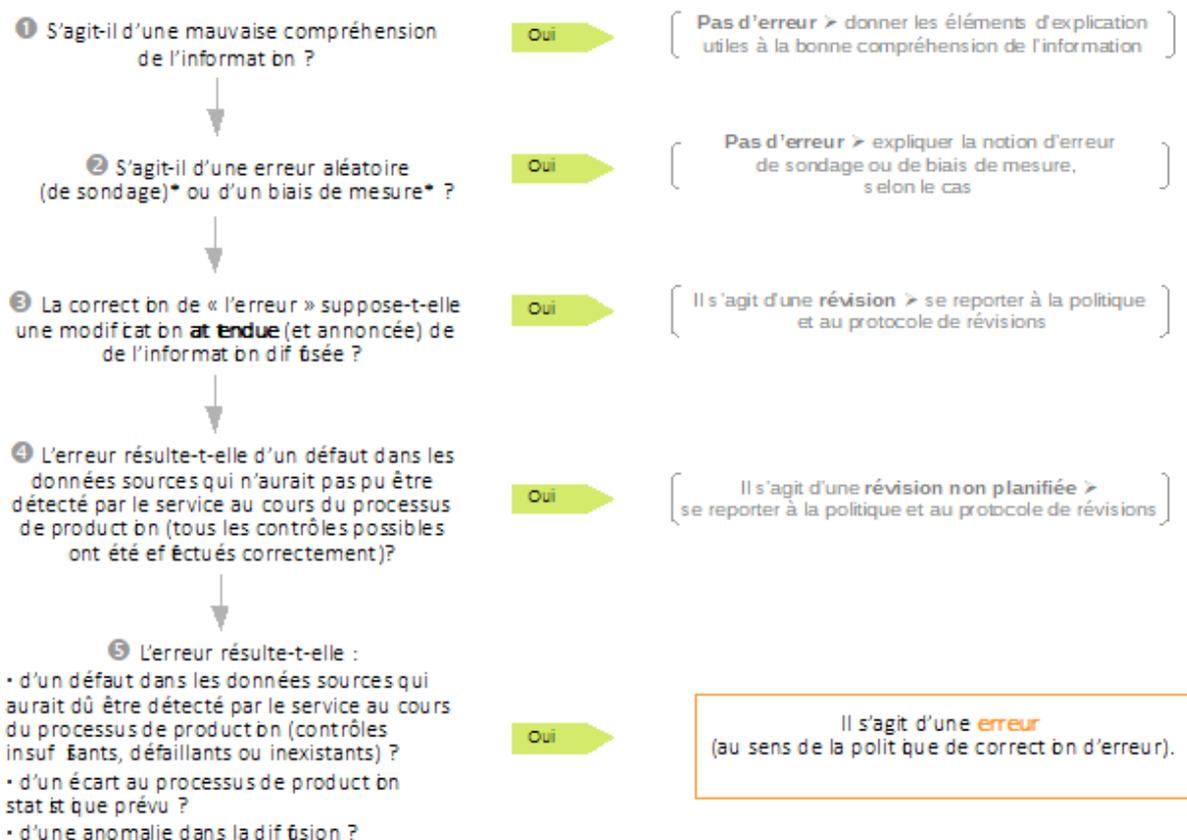
Le recensement des erreurs vise à :

- améliorer en continu les données sources, les processus et la qualité des résultats obtenus ;
- renseigner les utilisateurs sur les corrections effectuées, et plus généralement sur la qualité des statistiques.

L'ensemble des erreurs majeures répertoriées dans la production du service statistique public (Insee et services statistiques ministériels) fait chaque année l'objet d'une analyse présentée au Comité stratégique de la qualité du service statistique public, qui peut recommander toute disposition visant à améliorer la qualité des productions.

Les erreurs mineures ne font pas l'objet d'une documentation.

Annexe : arbre de décision pour identifier une erreur de production



* *Erreur liée au sondage* : incertitude liée au fait d'inférer de l'information sur une population à partir de celle collectée sur un échantillon de celle-ci (= précision).

* *Biais de mesure* : incertitude liée au défaut de couverture, à la non-réponse, aux hypothèses des modèles, etc.