

Qualitätsmerkmale für statistische Ergebnisse Definitionen, Maßzahlen, Erläuterungen

Die Bund-Länder-Arbeitsgruppe „Qualität der Statistikprodukte“ hat im Auftrag der Amtleiterkonferenz vom 12./13. November 2002 in Abstimmung mit dem Arbeitskreis für Fragen der mathematischen Methodik einen Bericht zur Qualität der Statistikprodukte erstellt. Dieser Bericht besteht aus zwei Teilen: den hier vorliegenden Definitionen, Maßzahlen und Erläuterungen zu Qualitätsmerkmalen für statistische Ergebnisse sowie „Empfehlungen für die Veröffentlichung von Qualitätsmerkmalen“.

Die Arbeitsgruppe konnte bei der Definition der Qualitätsmerkmale an Vorarbeiten anknüpfen wie die „Qualitätsstandards in der amtlichen Statistik“ vom November 2002. Sie hat sich intensiv mit dem Qualitätsbegriff, wie er im Europäischen Statistischen System erarbeitet wurde, auseinandergesetzt und die internationalen Standards für Qualität in der amtlichen Statistik einbezogen.

Die vorliegenden Definitionen stecken den allgemeinen Rahmen für Merkmale und Kriterien der Qualität statistischer Produkte ab. Sie können deshalb grundsätzlich für alle Fragen der Qualitätsbewertung in der amtlichen Statistik zu Grunde gelegt werden. Die Erläuterungen sind knapp gehalten und entsprechen weitgehend den auch im Europäischen Statistischen System verwendeten. Die vorgeschlagenen Maßzahlen sind jeweils am Ende der Definitionen und Erläuterungen übersichtlich zusammengefasst. Bei der Auflistung der möglichen Maßzahlen wurde keine Vollzähligkeit angestrebt, sondern es fand eine Vorauswahl im Hinblick auf die praktische Anwendbarkeit von Maßzahlen statt. Sie sind im Sinne eines Methodenbaukastens gedacht, aus denen die jeweils für besondere Statistiken sinnvollen Maße zusammengestellt werden können. Sie wurde nicht für eine vergleichende Bewertung oder Benchmarking von Statistiken bzw. Erhebungen konzipiert.

Diese Vorlage definiert also nicht die Qualität abschließend und schematisch für alle Statistiken. Vielmehr ist es erforderlich, dass die jeweils fachlich Verantwortlichen für jede einzelne Statistik unter Anwendung der hier gegebenen allgemeinen Definitionen und Maßzahlen diejenigen Spezifikationen vornehmen, die zur bestmöglichen Information für die Nutzer der Statistik führen.

Erst eine solche spezifizierte Beschreibung der Qualität einer Statistik könnte intern oder extern praktisch eingesetzt werden. Sie könnte beispielsweise zu einem internen Qualitätsbericht einer Statistik weiterentwickelt werden.

Bei einer externen Verwendung für die Nutzer der Statistik sind weitere Einschränkungen und Spezifikationen zu beachten, die nicht Gegenstand dieser Unterlage sind. Die Veröffentlichung von Qualitätsangaben hat eine vorrangige Bedeutung für die Nutzer der amtlichen Statistiken. Deshalb hat die Bund-Länder-Arbeitsgruppe „Empfehlungen für die Veröffentlichung von Qualitätsmerkmalen“ zusammengestellt und von den grundsätzlichen methodischen Fragen getrennt in einer eigenen Unterlage dargestellt. Diese Empfehlungen verstehen sich als Checkliste, die den Umfang der allgemein sinnvollen qualitätsbezogenen Angaben in Veröffentlichungen umschreibt und nicht als Mindestveröffentlichungsprogramm, das

schematisch angewendet werden könnte. Als praktische Anwendungsbeispiele wurden deshalb auch Spezifikationen für einzelne Statistiken beigefügt. Auch hier ist für die praktische Anwendung erforderlich, dass die fachlich Zuständigen für jede Statistik aufgrund der allgemeinen Checkliste der „Empfehlungen für die Veröffentlichung von Qualitätsmerkmalen“ die erforderlichen Spezifikationen vornehmen.

Im folgenden werden die Qualitätsmerkmale definiert, erläutert und geeignete Qualitätsmaße benannt.

Überblick: Für die statistischen Produkte der Statistischen Ämter des Bundes und der Länder gelten folgende Qualitätsmerkmale:

1. Relevanz,
2. Genauigkeit,
3. Aktualität und Pünktlichkeit,
4. Verfügbarkeit und Transparenz,
5. Vergleichbarkeit, z. B. von Länder- und Bundesergebnissen,
6. Kohärenz, z. B. von Monats- und Jahresergebnissen.

1. Relevanz

Statistische Informationen werden für den Bedarf der Nutzer produziert. Das Qualitätsmerkmal Relevanz erfüllen statistische Ergebnisse, wenn sie dem Bedarf der Nutzer entsprechen.

Die Relevanz statistischer Ergebnisse ist danach zu bewerten,

- welchen Bedarf die Nutzer haben,
- wie hoch der Grad der Bedarfsdeckung ist,
- wie breit das Spektrum der Nutzer ist,
- wie weit die statistischen Ergebnisse tatsächlich genutzt werden.

Unterschiedliche Klassen von Nutzern können vielfältige Erwartungen an die Qualität der statistischen Ergebnisse haben. Diese können z. B. im Hinblick auf Genauigkeit und Aktualität möglicherweise nicht für jeden Bedarf im gleichen Grad zu erfüllen sein. In der deutschen amtlichen Statistik ist auch der jeweilige Bedarf der spezifischen Benutzergruppen der Bundes- und Landesstatistik zu berücksichtigen.

Messgrößen für die Relevanz von Statistiken können erstellt werden, in dem die jeweiligen Nutzergruppen definiert und klassifiziert werden und ihr Bedarf hinsichtlich einer Reihe von Qualitätsdimensionen durch Befragungen oder andere Methoden festgestellt und ins Verhältnis zum Angebot und dessen tatsächlicher Nutzung gesetzt wird.

Informationen zum Bedarf von Nutzern entstehen regelmäßig auch auf dem Wege der Einbeziehung von Nutzern in die Entwicklung des statistischen Programms durch den Statistischen Beirat, Fachausschüsse und Äußerungen im Zuge von die Statistik betreffenden Gesetzgebungsverfahren. Erforderlichkeit und Umfang von statistischen Erhebungen sowie die Zweckbindung für die Nutzung durch Verwaltungen ergeben sich aus den die Statistik anordnenden Rechtsvorschriften. Die gesetzlichen bzw. rechtlichen Grundlagen von Statistiken werden in den allgemeinen Beschreibungsinformationen nachgewiesen.

Angaben, die zur Bewertung der Relevanz herangezogen werden können:

- Beschreibung und Klassifizierung der Nutzer,
- Beschreibung des Bedarfs der Nutzer (nach Nutzerklassen),
- Ergebnisse von Befragungen zur Nutzerzufriedenheit, insbesondere Bewertungen der Hauptnutzer; Hauptgründe für negative Einschätzungen,
- Vorgesehene Maßnahmen zur Erhöhung der Nutzerzufriedenheit,
- Verbreitung / Auflage der Publikationen mit den statistischen Ergebnissen; Web-Statistik der Zugriffe auf die statistischen Daten.

2. Genauigkeit

2. 1 Fehlertypologie für statistische Ergebnisse

Die Ergebnisse statistischer Erhebungen stimmen im Allgemeinen nicht exakt mit den wahren Werten der jeweiligen Grundgesamtheit überein. Die Abweichung zwischen dem Erhebungsergebnis und dem unbekanntem wahren Wert der Grundgesamtheit wird als **Gesamtfehler** des Ergebnisses bezeichnet.

Die **Genauigkeit** eines Erhebungsergebnisses wird vom Gesamtfehler bestimmt. Je größer der Gesamtfehler, desto geringer ist die Genauigkeit des Ergebnisses. Dementsprechend kann die Genauigkeit eines Ergebnisses definiert werden als Nähe des Ergebnisses zum wahren, aber unbekanntem Wert der Grundgesamtheit. Üblicherweise wird als Maß für die Genauigkeit der Gesamtfehler verwendet.

Zum Gesamtfehler tragen verschiedene Fehlerkomponenten bei, die sich in folgende Fehlerkategorien einteilen lassen:

(1) Stichprobenbedingte Fehler:

- Stichprobenzufallsfehler
- Stichprobenbedingte systematische Fehler (Verzerrungen, Bias)

(2) Nicht stichprobenbedingte Fehler:

- Fehler durch die Erfassungsgrundlage
- Messfehler
- Aufbereitungsfehler
- Fehler durch Antwortausfälle

Bei Vollerhebungen treten nur Fehler der Kategorie (2) auf, bei Stichproben Fehler der Kategorien (1) und (2). Bei den Fehlern der Kategorie (2) handelt es sich häufig um systematische Fehler, d.h. ihr Auftreten kann zu Verzerrungen der Ergebnisse führen.

Es ist üblich, als Maß für die Beurteilung des Gesamtfehlers eines Erhebungsergebnisses den mittleren quadratischen Gesamtfehler (Mean Square Error) zu verwenden, der sich additiv aus der Zufallskomponente und der systematischen Fehlerkomponente des Ergebnisses zusammengesetzt.

$$\text{MSE} = (\text{Standardfehler})^2 + (\text{Bias})^2$$

Der Standardfehler ist ein Maß zur Beurteilung des Stichprobenzufallsfehlers und gibt die sogenannte **Präzision** des Ergebnisses wieder.

Angaben, die zur Bewertung der Genauigkeit herangezogen werden können:

- Mittlerer quadratischer Gesamtfehler (Mean Square Error),
- absoluter oder relativer Standardfehler
- Größenordnung und Vorzeichen des Bias,
- Vergleich von Standardfehler und Bias,
- Beschreibung der Fehlerarten, die in die Schätzung des Gesamtfehlers eingeflossen sind,
- Qualitative Bewertung des Gesamtfehlers.

2. 2 Stichprobenbedingte Fehler

Bei stichprobenbedingten Fehlern unterscheidet man zwischen Stichprobenzufallsfehlern und systematischen Fehlern (Verzerrungen, Bias).

2.2.1 Stichprobenzufallsfehler

Stichprobenzufallsfehler sind nur für Ergebnisse von **Zufallsstichproben** definiert. Sie resultieren daher, dass nur ein Teil der Erhebungsgesamtheit befragt wird und die Zusammensetzung der Stichprobe vom Zufall abhängt. Auch wenn keine systematischen Fehler auftreten, können die Stichprobenergebnisse deshalb im Allgemeinen nicht exakt mit den Ergebnissen übereinstimmen, die mit einer gleichartigen Erhebung bei sämtlichen Einheiten der Erhebungsgesamtheit ermittelt worden wären, sondern sie weichen von diesen zufällig nach oben oder unten ab. Die Differenz ist der Stichprobenzufallsfehler.

Die Größe des Stichprobenzufallsfehlers hängt ab vom Stichprobenumfang, der Streuung (Varianz) des zu schätzenden Merkmalswertes, dem Stichprobendesign und dem Schätzverfahren für die Hochrechnung.

Die genaue Größe des Stichprobenzufallsfehlers ist ebenso wie seine Richtung im Einzelfall, d.h. für ein bestimmtes Stichprobenergebnis, unbekannt und kann auch nicht ermittelt werden, so dass die Zufallsabweichungen in den Stichprobenergebnissen rechnerisch nicht beseitigt werden können. Für Zufallsstichproben erlauben es die Gesetzmäßigkeiten der Wahrscheinlichkeitstheorie aber, anhand der Stichprobenwerte den Standardfehler eines Stichprobenergebnisses abzuschätzen und damit ein Intervall herzuleiten, in dem der interessierende Wert für die Erhebungsgesamtheit mit vorgebarbarer Wahrscheinlichkeit liegt.

Zur quantitativen Bewertung der Zufallsfehler von Stichprobenergebnissen kommen folgende Maßzahlen in Frage:

- Absoluter Standardfehler
- Relativer Standardfehler
- Konfidenzintervall.

Für allgemeine Veröffentlichungszwecke werden bevorzugt die geschätzten Standardfehler verwendet, da diese sich meistens auch zur näherungsweise Konstruktion von Konfidenzintervallen eignen. Für die Präzisionsabschätzung von Stichprobenergebnissen für Gesamt- und Mittelwerte wird in der Regel der relative Standardfehler (in der Literatur manchmal auch Variationskoeffizient der Schätzung genannt) benutzt. Er wird als Quotient aus dem absoluten Standardfehler (des Stichprobenergebnisses) und dem Stichprobenergebnis ermittelt. Bei aus der Stichprobe geschätzten Veränderungs- oder Anteilswerten ist die Angabe des absoluten Standardfehlers für den Nutzer meist leichter zu interpretieren.

Beide Maßzahlen definieren ein Intervall um das Stichprobenergebnis, in dem der zu schätzende Wert in der Regel mit ca. 68% Wahrscheinlichkeit liegt.

Für **Nicht-Zufallsstichproben**, d.h. Teilerhebungen, denen ein „bewusstes“ oder „gezieltes“ Auswahlverfahren zugrunde liegt (z.B. typische Auswahl, Quotenverfahren, Abschneideverfahren), ist es theoretisch nicht möglich, Stichprobenzufallsfehler zu schätzen. Wenn aber unterstellt werden kann, dass das Auswahlprinzip einer solchen Stichprobe der Zufallsauswahl nahe kommt, so ist es üblich, die Fehlermaße der Zufallsstichproben als Näherungslösung anzuwenden.

2.2.2 Stichprobenbedingte systematische Fehler (Verzerrungen, Bias)

Stichprobenbedingte systematische Fehler sind auf Ursachen zurückzuführen, die nur bei Stichprobenerhebungen auftreten.

Zu dieser Fehlerkategorie gehören:

- Verzerrungen durch die Auswahlmethode und –technik, wenn kein echtes Zufallsverfahren angewendet wird;
- Verzerrungen durch das Hochrechnungsverfahren, wenn z.B. eine gebundene Hochrechnung (Regressionsschätzung, Verhältnisschätzung) angewendet wird oder eine Anpassung (Kalibrierung) an Eckwerte aus anderen Quellen durchgeführt wird.

Generell gilt, dass die stichprobenbedingten systematischen Fehler nicht aus der Stichprobe selbst abschätzbar sind. Lediglich der Bias, der durch die Anwendung gebundener Hochrechnungsverfahren verursacht wird, lässt sich in der Regel modellbezogen abschätzen. Er wird bei größeren Stichprobenumfängen in der Regel vernachlässigbar klein.

Angaben, die zur Bewertung stichprobenbedingter Fehler herangezogen werden können:

- Auswahlmethode und –technik,
- Schätzverfahren für die Hochrechnung (verzerrt/unverzerrt),
- Schätzwerte für die absoluten und relativen Standardfehler,
- geschätzte Konfidenzintervalle,
- Größe des Bias, der durch verzerrende Hochrechnungsverfahren verursacht wird,
- Methode, die zur Fehlerabschätzung verwendet worden ist. Hinweise, ob die Auswirkungen von Nonresponse, Imputationen o.ä. bei der Fehlerabschätzung berücksichtigt worden sind.
- Bezeichnung der Software für die Fehlerabschätzung (Standard-/Individualsoftware),
- bei Nichtzufallsstichproben: Aussage, in wie weit die Fehlermaße näherungsweise angewendet werden können.

2. 3 Nicht stichprobenbedingte Fehler

Nicht stichprobenbedingte Fehler betreffen Stichproben- und Vollerhebungen gleichermaßen und treten in allen Phasen des Datenerhebungs- und Aufbereitungsprozesses auf.

2. 3.1 Fehler durch die Erfassungsgrundlage

Zu Fehlern durch die Erfassungsgrundlage kommt es, wenn in der Erfassungsgrundlage (z.B. Register, Anschriftendatei) einer Erhebung

- nicht alle Einheiten der Grundgesamtheit enthalten sind,
- Einheiten der Grundgesamtheit mehrfach vertreten sind,
- Einheiten enthalten sind, die nicht zur Grundgesamtheit gehören.

Als Grundgesamtheit wird dabei die Gesamtheit der Einheiten bezeichnet, über die Aussagen gemacht werden sollen. Die Erfassungsgrundlage ist die Basis sowohl für die Durchführung von Voll- als auch von Stichprobenerhebungen. Bei Stichprobenerhebungen stellt sie die Auswahlgrundlage dar. Während die Grundgesamtheit laufend zeitlichen Änderungen ihrer Zusammensetzung unterworfen ist, wird die Erfassungsgrundlage zu einem bestimmten Zeitpunkt fixiert und kann in der Folgezeit veralten. Die Folgen können Übererfassungen, Untererfassungen, mehrfach vertretene Einheiten sowie – falls die Erfassungsgrundlage Klassifikationen enthält, die einem Wandel unterliegen können – Fehlklassifikationen sein.

2.3.2 Messfehler

Messfehler sind Fehler, die in der Datenerhebungsphase auftreten. Ihre Ursachen können wie folgt kategorisiert werden:

Erhebungsformular (Fragebogen): z.B. missverständliche Formulierung der Fragen, unübersichtliche Fragebogengestaltung;

Interviewer: Antwortbeeinflussung durch den Interviewer;

Befragte: Bewusste oder unbewusste Antwortfehler.

Messfehler verursachen Verzerrungen der Ergebnisse und können bei Stichproben auch die Schätzung des Zufallsfehlers beeinflussen. Um durch Erhebungsformulare oder Interviewer verursachte Messfehler abschätzen zu können, sind Wiederholungsbefragungen mit unterschiedlichen Fragebögen und Interviewern erforderlich. Antwortfehler können im Allgemeinen nicht abgeschätzt werden, da dazu Informationen aus anderen Quellen über die Befragten verfügbar sein müssten.

Ein Teil der Messfehler wird in der Aufbereitungsphase bei der Plausibilisierung der Angaben berichtigt.

2.3.3 Aufbereitungsfehler

Aufbereitungsfehler sind Fehler, die in der Phase der Aufbereitung von Statistiken entstehen, z.B. bei der **Verschlüsselung** und **Signierung**, der **Datenerfassung**, bei Korrekturen im Rahmen der **Plausibilisierung** der Angaben. Auf der anderen Seite können in dieser Phase Messfehler, z.B. Eintragungsfehler, auch wieder berichtigt werden.

Die Art der Korrektur bei der Plausibilisierung kann die Ergebnisse beeinflussen, wenn gezielt bestimmte Formen der Fehlerauflösung bevorzugt werden. Dies kann zu Verzerrungen führen und bei Stichproben auch die Zufallsfehlerschätzung beeinflussen.

Der Anteil der Korrekturen kann für jedes Merkmal grundsätzlich mit der **Korrekturquote** gemessen werden. Die Korrekturquote ist definiert als Anteil der erhobenen Einheiten mit korrigierten oder eingesetzten Merkmalsausprägungen an den erhobenen Einheiten insgesamt.

2.3.4 Fehler durch Antwortausfälle (Nonresponse)

Zu Fehlern durch Antwortausfälle kommt es, wenn es – gleich aus welchen Gründen – nicht gelingt, die vollständigen Informationen zu allen Einheiten oder für alle Variablen einer Stichprobe oder Vollerhebung zu sammeln. Antwortausfälle können zu Verzerrungen der Ergebnisse führen und bei Stichproben den Zufallsfehler erhöhen.

Es werden zwei Arten von Antwortausfällen unterschieden:

- **Unit-Nonresponse:** eine Einheit liefert überhaupt keine Daten
- **Item-Nonresponse:** die Daten einer Einheit sind zu bestimmten Merkmalen unvollständig.

Die Annahmen zum Verhalten von Nichtauskunftgebenden werden in der Literatur je nach Ausfall-Mechanismus drei Kategorien zugeordnet:

- **Rein zufällige Datenausfälle** (missing completely at random – MCAR): zwischen Nichtauskunftgebenden und Auskunftgebenden besteht kein Unterschied.
- **Zufällige Datenausfälle** (missing at random – MAR): die Datenausfälle hängen nicht von den fehlenden Merkmalswerten ab, aber von Hilfsmerkmalen.
- **Nicht zufällige (systematische) Datenausfälle** (not missing at random – NMAR): Die Datenausfälle hängen ab von den fehlenden Merkmalswerten. Der Mechanismus des Datenausfalls heißt dann „informativ“. So werden beispielsweise Antworten auf

Fragen nach dem Einkommen bei höheren Werten dieses Merkmals häufiger verweigert.

Bei nicht zufälligen systematischen Datenausfällen sind größere Verzerrungen zu erwarten als bei zufälligen Datenausfällen. Rein zufällige Datenausfälle führen nicht zu Verzerrungen. Die Zuordnung der Ausfälle zu diesen Kategorien dürfte im Einzelfall allerdings häufig schwierig sein.

Das Ausmaß von Antwortfehlern kann durch **Antwortquoten** gemessen werden. Entsprechend den zwei Arten von Antwortausfällen werden zwei Antwortquoten unterschieden:

- Die **Antwortquote bei Erhebungseinheiten** ist definiert als Anteil der erhobenen Einheiten an den Erhebungseinheiten insgesamt.
- Die **Antwortquote bei Merkmalen** kann sowohl als Anteil der erhobenen Einheiten mit Angaben zu dem Merkmal an den erhobenen Einheiten insgesamt als auch an der Gesamtheit der Erhebungseinheiten definiert werden. Die jeweilige Definition ist bei der Interpretation des Ergebnisses zu beachten.

Bei der Interpretation der Antwortquoten ist zwischen freiwilligen Erhebungen und Erhebungen mit Auskunftspflicht zu unterscheiden. Die Antwortquote bei Erhebungseinheiten zeigt die Beteiligung an, macht aber keine zuverlässige Aussage zur Repräsentativität. Oftmals werden **Imputationsverfahren** eingesetzt, um fehlende Daten zu ersetzen, z.B. Ersatzvornahmen aus vorhandenen Registern, Doppelung vergleichbarer Fälle, Einsetzung von Mittelwerten. Hierbei ist zu beachten, dass sich Imputationen auf die Genauigkeit der Ergebnisse auswirken; einige der Verfahren führen z.B. zu einer fälschlichen Unterschätzung des Zufallsfehlers.

Angaben, die zur Bewertung nicht stichprobenbedingter Fehler herangezogen werden können:

Zu Fehlern durch die Erfassungsgrundlage

- Schätzung der Quoten für Übererfassungen, Untererfassungen, mehrfach vertretene Einheiten und Fehlklassifikationen,
- Informationen zur Aktualität der Erfassungsgrundlage,
- Bewertung der Verzerrung durch Fehler, die durch die Erfassungsgrundlage verursacht worden sind.

Zu Messfehlern

- Erhebungsverfahren,
- Ergebnisse von Kontrollerhebungen,
- Ergebnisse des Fragebogen-Pretests,
- Qualität der Interviewer,
- Korrekturquote bei der Plausibilisierung,
- Bewertung der Verzerrung durch Messfehler.

Zu Aufbereitungsfehlern

- Aufbereitungsverfahren,
- Fehlerquote bei Signierung und Datenerfassung,
- Beschreibung des Plausibilisierungsverfahrens,
- Informationen über die verwendeten Imputationsverfahren,

- Korrekturquote,
- Bewertung der Verzerrung durch Aufbereitungsfehler.

Zu Fehlern durch Antwortausfälle

- Antwortquoten bei Einheiten und Merkmalen,
- Nonresponse-Quoten gegliedert nach Ausfallursachen,
- Erkenntnisse zum Ausfall-Mechanismus,
- Beschreibung der Imputationsverfahren und Bewertung ihrer Auswirkungen auf die Schätzung der Ergebnisse und Standardfehler,
- Bewertung der Verzerrung durch Antwortausfälle.

Qualität und Vollständigkeit der Dokumentationen über Erhebung und Aufbereitung

2.3.5 Genauigkeit vorläufiger Daten

Einen Sonderfall des Umgangs mit fehlenden Angaben stellt die Veröffentlichung von **vorläufigen Ergebnissen** dar. Insbesondere bei kurzfristigen Statistiken ist es meist üblich, Einheiten, deren Angaben verspätet eintreffen, zunächst als Antwortausfälle zu behandeln und vorläufige Ergebnisse zu erstellen. Der Effekt solcher Antwortausfälle lässt sich durch einen a posteriori – Vergleich von vorläufigen und endgültigen Ergebnissen ermitteln. Ein geeigneter Indikator dafür ist der **Revisionsumfang**, d.h. die Abweichung zwischen vorläufigem und endgültigem Ergebnis. Aus einer genügend großen Zahl von Revisionen zu einer kurzfristigen Statistik kann der **Revisionsbedarf**, d.h. der zu erwartende Korrekturumfang vorläufiger Ergebnisse dieser Statistik, abgeschätzt werden.

Zur Abschätzung des Revisionsbedarfs vorläufiger Ergebnisse ist insbesondere der arithmetische Mittelwert der in der Vergangenheit beobachteten Abweichungen zwischen vorläufigen und endgültigen Werten ohne Vorzeichenberücksichtigung geeignet. Eine weitere sinnvolle Maßzahl ist der arithmetische Mittelwert der in der Vergangenheit beobachteten Revisionen mit Vorzeichenberücksichtigung. Diese Maßzahl kann einen Hinweis auf einen möglichen systematischen Fehler des vorläufigen Ergebnisses liefern.

Als alternative Maßzahlen könnten auch die Mediane der Abweichungen verwendet werden. Der Vorteil liegt in der größeren Robustheit, d.h. der geringeren Empfindlichkeit gegenüber Ausreißern. Ergänzend könnte auch die Streubreite der Differenzen dargestellt werden, entweder in Form der Standardabweichung oder eines Quantilabstands, z.B. der Mittelbereich der Abweichungen, in dem 2/3 der Fälle liegen.

Angaben, die zur Bewertung vorläufiger Ergebnisse herangezogen werden können:

- Anteil der einbezogenen Einheiten und Merkmalswerte, ggf. nach Klassen gegliedert, die zum vorläufigen Ergebnis beigetragen haben,
- arithmetischer Mittelwert der in der Vergangenheit beobachteten Revisionen mit und ohne Vorzeichenberücksichtigung,
- Median der in der Vergangenheit beobachteten Revisionen,
- Standardabweichung oder Quantilsabstand der in der Vergangenheit beobachteten Revisionen,
- Gründe für Revisionen.

3. Aktualität und Pünktlichkeit

Aktualität und Pünktlichkeit beziehen sich in unterschiedlicher Weise auf Zeitpunkte und Termine:

Aktualität einer Statistik bezieht sich auf die Zeitdifferenz zwischen dem Berichtszeitpunkt bzw. dem Ende des Berichtszeitraums und dem Zeitpunkt, zu dem die statistische Information planmäßig oder regelmäßig für die Nutzer verfügbar wird. Bei der Veröffentlichung von vorläufigen Ergebnissen kann dem Kriterium der Aktualität ein höheres Gewicht als dem Kriterium der Genauigkeit zukommen.

Der Zeitpunkt und die Art der Veröffentlichung sind zu spezifizieren. Es kann sich um die Veröffentlichung vorläufiger oder endgültiger Ergebnisse handeln, von Eckzahlen oder tief gegliederten Tabellen, in elektronischer oder gedruckter Form usw.

Die Aktualität wird gemessen als Zeitspanne zwischen dem Berichtszeitpunkt bzw. dem Ende des Berichtszeitraums und den planmäßigen bzw. regelmäßigen Veröffentlichungsterminen vorläufiger bzw. endgültiger Ergebnisse. Die Angabe der Aktualität ist ein Qualitätsmerkmal innerhalb der Metainformationen für eine Statistik und wird als allgemeine Beschreibungsinformation (Zeitangabe) dargestellt.

Als **Pünktlichkeit** wird die Einhaltung der vorab festgelegten Veröffentlichungstermine bezeichnet. Um alle Nutzer hinsichtlich der Verfügung über die Ergebnisse gleich zu behandeln, sollten Veröffentlichungstermine vorab festgelegt und den Nutzern bekannt gemacht werden.

Hinsichtlich der Produktion von Statistiken ist das Kriterium der Pünktlichkeit anzuwenden. Die Pünktlichkeit einer Statistik bezieht sich auf die Einhaltung bzw. Über- oder Unterschreitung der geplanten Produktions- und Veröffentlichungstermine. Auch bei vorfristiger Veröffentlichung von statistischen Ergebnissen oder Teilergebnissen wird das Kriterium der Pünktlichkeit nicht eingehalten.

Aktualität und Pünktlichkeit als Qualitätsmerkmale für statistische Ergebnisse sind auf die Nutzer der Statistik bezogen. Die in Arbeits- und Zeitplänen niedergelegte Terminüberwachung der Statistikproduktion ist hingegen eine Funktion der internen Qualitätskontrolle.

Angaben, die zur Bewertung der Aktualität und Pünktlichkeit herangezogen werden können

- Zeitspanne zwischen Berichtszeitpunkt (bzw. Ende des Berichtszeitraums) und Veröffentlichungstermin der vorläufigen und der endgültigen Ergebnisse,
- Zeitspanne zwischen dem planmäßigen und dem tatsächlichen Veröffentlichungstermin der Ergebnisse,
- Gründe für verspätete Veröffentlichungen,
- Vorgesehene Maßnahmen zur Verbesserung der Aktualität bzw. der Pünktlichkeit.

4. Verfügbarkeit und Transparenz

Verfügbarkeit und Transparenz beziehen sich darauf, ob die Nutzer einfach und schnell auf statistische Daten zugreifen können, diese innerhalb einer annehmbaren Frist nach üblichen Standards und in der erwarteten Form erhalten und dabei hinsichtlich Konzept und Methoden angemessen informiert und unterstützt werden. Hierfür müssen ihnen sowohl die Zugangs- und Bezugsbedingungen als auch die erforderlichen Auskünfte über das Datenangebot und die Beschreibungsinformationen und Metadaten zugänglich und transparent sein.

Verfügbarkeit bezieht sich auf die praktische Zugänglichkeit der Daten: an wen sich der Nutzer wenden muss; wie er bestellen kann; die Lieferzeit; die Gestaltung der Preispolitik; die Vertriebsbedingungen; die Verfügbarkeit von Makro- und Mikrodaten; unterschiedliche Formate und Medien wie Dateien, CD-Rom, Internet, Papiertabellen usw. , weitergehende Serviceleistungen.

Die Verfügbarkeit von statistischen Daten wird grundsätzlich durch die statistische Geheimhaltung begrenzt.

Transparenz bezieht sich auf das informationelle Umfeld der Daten: Zum einen die Ergänzung der statistischen Informationen durch Erläuterungen und Metadaten; methodische Informationen sowie Informationen über die Datenqualität. Zum anderen weiterführende und interpretierende Textinformationen; Abbildungen; Karten usw.

Das informationelle Umfeld der Daten ist für die Nutzer transparent, wenn sie leichten Zugang zu allen erforderlichen, die Statistik beschreibenden Metadaten haben. Dies ist insbesondere dann gegeben, wenn die statistischen Produkte sowohl statistische Ergebnisse als auch vollständige Metadaten in sich vereinigen sowie Referenzen zu den wichtigsten weiterführenden Informationen enthalten.

Angaben, die zur Bewertung der Verfügbarkeit und Transparenz herangezogen werden können:

- Beschreibung der Bedingungen, unter denen die Daten zugänglich sind: u.a. Bezugsadresse, Vertriebsbedingungen, Kontaktinformationen, Kosten,
- Ergänzende Informationen zu den Daten: u.a. Methoden, Konzepte, Genauigkeit, weiterführende Veröffentlichungen,
- Auflistung der Veröffentlichungsmedien.

5. Vergleichbarkeit von Statistiken

Statistische sollen zuverlässige zeitliche, räumliche und fachliche Vergleiche ermöglichen, auch wenn sie sich auf unterschiedliche Grundgesamtheiten beziehen. Die Vergleiche können sich auf internationale oder regionale, zeitliche und kategoriale Unterschiede erstrecken. Abweichungen können durch unterschiedliche Erhebungskonzepte und Messverfahren auftreten. In der föderal organisierten Statistik steht die zuverlässige zeitliche und räumliche Vergleichbarkeit von Länder- und Bundesergebnissen an erster Stelle. Die einzelnen Schritte der Datenproduktion werden daher möglichst einheitlich gestaltet. Auf die Kontinuität der Ergebnisse wird geachtet.

Die Statistik versucht den unbekanntem wahren Wert von Merkmalen der Grundgesamtheit zu messen. Diese Merkmale müssen vergleichbar sein. Deshalb sollten relative Abweichungen/Unterschiede wie beispielsweise die folgenden betrachtet werden:

- Abweichungen bei Erhebungskonzepten, Grundgesamtheit, Erfassungsgrundlagen,
- Abweichungen bei Definitionen sowie Mess- und Schätzverfahren,
- Abweichungen der gesetzlichen Vorschriften,
- Abweichungen in den Klassifikationen und
- Abweichungen in der regionalen Gliederung.

Vergleichbarkeit kann aus folgenden Perspektiven beurteilt werden:

- Räumliche Vergleichbarkeit
- Zeitliche Vergleichbarkeit
- Vergleichbarkeit zwischen Bereichen

Räumliche Vergleichbarkeit:

Es ist ein Vergleich der Landes- und Bundesergebnisse vorzunehmen, falls bei den Ergebnissen Unterschiede vor allem bei der Definition der statistischen Einheiten, der Bezugsgesamtheit, den Klassifikationen und Definitionen der Variablen und in den Erhebungskonzepten der Regionen bestehen. Nach Möglichkeit sind die Unterschiede zu quantifizieren.

Zeitliche Vergleichbarkeit:

Eventuelle Brüche in Zeitreihen sind zu beschreiben. Unterschiede in Konzept oder Methoden zwischen der letzten und der vorhergehenden Periode sind darzustellen z.B. Veränderung der Klassifikationen, der Definitionen, der statistischen Methoden, der Grundgesamtheit bzw. des Erfassungsbereichs, den Datenverarbeitungsmethoden. Die Größe der Effekte ist, wenn möglich, quantitativ einzuschätzen.

Vergleichbarkeit zwischen Bereichen:

Statistiken werden meist auf der Grundlage von Bereichen, die durch Klassifikationen definiert sind, verglichen. Hierzu zählen Klassifikationen zur Gliederung wirtschaftlicher Aktivitäten, von Waren und Dienstleistungen, sozioökonomischen Merkmalen, Regionalangaben usw. An Bedeutung zugenommen haben die Anforderungen an die Vergleichbarkeit mit internationalen Klassifikationen, z.T. ist selbst ihre unmittelbare Anwendung für nationale Statistiken erforderlich.

Auch konzeptionelle Unterschiede von Statistiken sind zu beachten. Sie betreffen überwiegend die Definition des Merkmals, den Bezugszeitraum, die Definition der statistischen Einheit und das statistische Maß.

Angaben, die zur Bewertung der Vergleichbarkeit herangezogen werden können:

Erhebungskonzepte

- Erhebungsmerkmale,
- Statistische Maße (Indikatoren),
- Erhebungseinheiten,
- Grundgesamtheit,
- Bezugszeitraum,
- Untersuchungsbereiche,
- Vorschriften / Normen,
- Für einen Untersuchungsbereich spezifische konzeptionelle Aspekte.

Messverfahren

- Stichprobendesign,
- Stichprobenfehler,
- Erhebungsverfahren,
- Aufbereitungsverfahren,
- Hochrechnungsverfahren.

sonstige Unterschiede

- Abweichungen bei Definitionen und Klassifikationen,
- Unterschiede der regionalen Gliederung,
- Abweichungen in den Zeitreihen,
- Gründe für Revisionen,
- ergänzende Informationen zu den Daten,

6. Kohärenz der Ergebnisse

Statistische Ergebnisse über gleiche Sachverhalte der gleichen Grundgesamtheit sollen kohärent sein, d. h. konsistent und widerspruchsfrei sein. Kohärenz bezieht sich beispielsweise auf den Grad, zu dem die Ergebnisse von Monats- und Jahreserhebung der selben Statistik übereinstimmende Ergebnisse liefern. Abweichungen können sich aus unterschiedlichen Erhebungs- und Messverfahren oder unterschiedlicher Genauigkeit ergeben. Unterschiede der Ergebnisse von Erhebungen über gleiche Sachverhalte (z. B. Ergebnisse aus Erhebungen unterschiedlicher Periodizität, vorläufige und endgültige Ergebnisse, Ergebnisse aus Erhebungen mit unterschiedlichen Quellen) sollten quantifiziert und erläutert werden, um es dem Nutzer zu ermöglichen, diese untereinander in Beziehung zu setzen.

Angaben, die zur Bewertung der Kohärenz herangezogen werden können:

Kohärenz zwischen vorläufigen und endgültigen Statistiken

- Vergleich der Ergebnisse,
- Bewertung der Auswirkung der Fehler der wichtigsten Merkmale (Stichprobenbedingte Fehler, Fehler durch die Erfassungsgrundlage, Messfehler, Aufbereitungsfehler und Fehler durch Antwortausfälle).

Kohärenz jährlicher und unterjährlicher Statistiken

- Jährlicher Vergleich der Statistiken und ggf. der Veränderungsraten unter Berücksichtigung der Genauigkeit, die für beide Arten von Statistiken geschätzt werden sollte,
- ggf. Bewertung von Unterschieden in den jeweiligen Erhebungskonzepten.

Kohärenz von Statistiken in denselben Bereichen

- Jährliche Abweichungen in gemeinsamen Merkmalen unter Berücksichtigung der Genauigkeit und von Unterschieden in den jeweiligen Erhebungskonzepten,
- falls vorhanden, Übersicht über Spiegelstatistiken,
- Schätzung der durch Unterschiede hinsichtlich Erhebungskonzepten und Genauigkeit bedingten Asymmetrien.

Kohärenz mit den Volkswirtschaftlichen Gesamtrechnungen

- Vergleich der Ergebnisse zu gemeinsamen Merkmalen.