

Stand: 12.10.2010

PIAFStat-Verfahren für die „Hohenheim-Gülzower Serienauswertung“

Name: PHI

Label: optimale Datentransformation

Macros: %boxcox

Funktionen: PFAD; TYP; MEAN; DATEN1; ADD; DUBL_CHECK; ZWEIDEUTIG1
OPTION2; MC_BOX

Konzeption: Michel, Piepho

Umsetzung SAS, PIAFStat: Zenk

SAS-Macros: Möhring

Seit Version vom 10.3.2008 wurden insbesondere folgende Neuheiten in das Programm eingefügt:

- Erweiterung der Option 04 – mehrfaktoriell (j/n) um Stufenmittel und Behandlungseffekt
- Nutzung von Erfahrungswerten als Startwerte für Phi (externe Datei PHISart.xls)

1. Ziel und Leistung des Verfahrens

- berücksichtigt die Modellvoraussetzungen, insbesondere den Komplex „Additivität, Varianzhomogenität und Normalverteilung“ in der Weise, dass gegebene Daten durch Transformation optimal an die Modelle angepasst werden können und damit die größte Plausibilität aufweisen
- dafür wird der optimale Transformationsparameter (ϕ bzw. φ) geschätzt, der in den Verfahren ‚VK‘ und ‚MW‘ eingesetzt werden kann und zur optimierten, unverzerrten Mittelwertschätzung beiträgt

Aufgrund des relativ großen Rechenzeitbedarfs (iteratives Verfahren) wird vorgeschlagen:

- Bestimmung des optimalen Transformationsparameter ϕ in größeren Abständen
Phi kann, wie auch die Varianzkomponenten, zeitlich unabhängig von der eigentlichen Auswertung (Verfahren ‚MW‘) bestimmt werden.
- Nutzung der Option: Reduktion des Datensatzes / Versuche je Sorte (z.B. auf 8 setzen)
- Nutzung der Vorgabe geeigneter Startwerte zur Beschleunigung der Iteration
In der Option 013 ‚Berechnung des optimalen Transformationsparameters‘ ist optional die Angabe zweier Startwerte möglich. Als default werden hier die Werte 0 (Log-Transformation) und

1 (keine Transformation) verwendet. Für die Eingabe der Startwerte ist zu beachten, dass das Optimum möglichst zwischen den Startwerten liegen sollte.

2. IT –Voraussetzungen und –Hinweise (Dr. Zenk)

Zu den allgemeinen Hinweisen siehe Info-VK!

Leistung der Funktionen:

Funktion	Leistung
PFAD	Pfade festlegen
TYP	Automatische Generierung des Typs im Datensatz
MEAN	Komprimierung doppelter Datensätze (z.B. bei Intensitätsstufe II und III)
DATEN1	Vorbereitung des Datensatzes, Datenbasis, und deren Kontrolle
OPTION2	gewählte Optionen werden im Output ausgegeben
ADD	Automatische Generierung der Additionskonstante
DUBL_CHECK	Prüfung des Datensatzes auf Dubletten
ZWEIDEUTIG1	Prüfung der BSA-Kennnr. und der Sortennamen auf Eindeutigkeit
MC_BOX	Macro-Aufruf zur Bestimmung des optimalen Phi

3. Optionen

- O0: Kulturart: siehe ‚MW‘ bzw. ‚VK‘
- O1: Pfad zu den Ergebnis - Dateien: siehe ‚MW‘ bzw. ‚VK‘
- O2: Pfad zu den externen Dateien: siehe ‚MW‘ bzw. ‚VK‘
- O3: Datenbasis: siehe ‚MW‘ bzw. ‚VK‘
- O4: mehrfaktoriell (ja/nein): siehe ‚MW‘ bzw. ‚VK‘
- O5: Reduktion des Datensatzes: siehe ‚MW‘ bzw. ‚VK‘
- O8: Regionen: siehe ‚MW‘ bzw. ‚VK‘
- O9: Sorten-Gruppen: siehe ‚MW‘ bzw. ‚VK‘
- O10: Datensatz abspeichern: siehe ‚MW‘ bzw. ‚VK‘
- O11: Übersicht und Kontrolle: siehe ‚MW‘ bzw. ‚VK‘
- O12: Dubletten-Check: siehe ‚MW‘ bzw. ‚VK‘

- O13: Berechnung des optimalen Transformationsparameters:

Die Option sollte im Regelfall als zentrales Ziel des Verfahrens auf ‚ON‘ gesetzt sein. Ein Schätzwert von $\phi=1$ bedeutet, dass kein Transformationserfordernis besteht. Im Zufallsbereich wird es immer Abweichungen von 1 geben, die vorrangig erst bei signifikanter Abweichung berücksichtigt werden sollten. Signifikanz kann mit nachfolgender Option geprüft werden. (siehe auch methodische Hintergründe)

Innerhalb dieser Option ist das Verändern der Vorgabewerte für Minimum (default= 0) und Maximum (default= 1) möglich. Wie unter 1. erläutert, wird die Nutzung dieser Möglichkeit empfohlen, um die Rechenzeit zu optimieren. Für die Eingabe der Startwerte ist zu beachten, dass das Optimum möglichst zwischen den Startwerten liegen sollte. Unsere Erfahrungswerte liegen in folgenden Bereichen:

Erträge	0,7 - 1,2
Behandlungseffekte	0,2 - 0,8
Prozentwerte	0,0 - 0,6.

Der ermittelte optimale Transformationsparameter wird im List-Fenster ausgegeben und zudem in einer Access-Datei [Kulturart].mdb als Tabelle ‚phi[A1]‘ abgespeichert.

- O14: Ausgabe der Graphik:

Die Grafik zeigt die Güte des Modells (Log-Likelihood) in Abhängigkeit vom Transformationsparameter. Sie dient der Signifikanzprüfung und Ausweisung eines Vertrauensintervalls für den geschätzten Parameter.

Die Grafik wird am Bildschirm angezeigt, nach Enter wird das Programm weiter abgearbeitet.

Die Ausgabe erzeugt zudem eine Datei ‚phi[A1].rtf‘, die die Grafik enthält.

4. Standard-Output

O0 bis O12: siehe Verfahren ‚VK‘ und ‚MW‘

Als Ausgabe wird der optimale Parameter ausgegeben und eine Graphik erstellt. Diese zeigt die Güte des Modells (als Log-Likelihood angegeben) in Abhängigkeit vom Transformationsparameter. Zur Signifikanzprüfung und Ausweisung eines Vertrauensintervalls für den geschätzten Parameter wird eine Linie angezeigt. Alle Transformationsparameterwerte, deren Log-Likelihood größer als die eingezeichnete Linie sind, unterscheiden sich nicht signifikant vom optimalen Transformationsparameter. Interessant ist insbesondere, ob der Wert 1 für den Transformationsparameter (gleich untransformiert) sich vom optimalen Wert unterscheidet.

5. Methodische Hintergründe und Diskussion (MICHEL)

Modellannahmen:

- Additivität
- Varianzhomogenität
- Normalverteilung
- Unabhängigkeit
- missing (*completely*) at random (zufälliges Fehlen; MAR bzw. MCAR)
diese Modellvoraussetzung wurde in der „Piepho-Studie“ untersucht mit dem Ergebnis, dass sie i.d.R. als hinreichend erfüllt angesehen werden kann, wenn keine willkürliche Dateneinschränkung vor der Verrechnung erfolgt

→ Dieser Gesamtkomplex kann mit dem Verfahren ‚phi‘ vorab geprüft und im Ergebnis durch Verwendung eines vom Verfahren geschätzten Transformationsparameters ‚phi‘ berücksichtigt werden.

Die Modellannahmen können leider nicht voneinander isoliert geprüft werden, da sie einen abhängigen Komplex darstellen. Wenn z.B. Additivität als ursächliche Komponente verletzt ist, so ist automatisch auch Varianzhomogenität und Normalverteilung nicht gegeben. Es wäre zwar wünschenswert, die ursächlich verletzte Annahme festzustellen, dies ist aber nicht ohne weiteres möglich.

Die Transformation garantiert hier keine Erfüllung aller Voraussetzungen. Es handelt sich lediglich um eine Schätzung, welche die größte Likelihood (Mutmaßlichkeit) und damit die größte Plausibilität aufweist.

In den Serienauswertungen der LDS wird die Erfüllung der Modellvoraussetzungen derzeit m.W. i.d.R. nicht geprüft, sondern als gegeben vorausgesetzt. Dieses Defizit soll hiermit behoben werden und gleichzeitig soll durch Transformation eine optimale Auswertung ermöglicht werden.

Warum Additivität besonders hinterfragt ?

Mit der Ausweitung der Auswertungsräume durch das Überlappungsprinzip geht eine zunehmende Ertragsdifferenziertheit der Versuchsdurchschnittserträge einher. Dies hatte mich bewogen, a) die Richtigkeit der Additivitätsannahme im Allgemeinen zu hinterfragen und ein Werkzeug zu schaffen, mit dem in konkreten Serien-Auswertungen Abweichungen ggf. berücksichtigt werden können. Bisher wurde die Frage vernachlässigt, ob das Datenverhalten überwiegend additiv oder multiplikativ ist – es wird ungeprüft nach additivem Modell gerechnet und dazu im Widerspruch ebenso ungeprüft nach multiplikativem Modell dargestellt (Relativwerte). Die Darstellung von Relativzahlen erfolgt nicht wegen

der Unterstellung eines multiplikativen Modells, sondern wegen der erleichterten intuitiven Tabelleninterpretation (z.B. bei der Darstellung von Sorten-Erträgen in verschiedenen Versuchen mit unterschiedlichem Durchschnittsertrag) und wegen der Gewöhnung an eine einheitliche Zahlendimension um 100 % unabhängig von der Mengeneinheit (dt, MJ, Anzahl). Dieser Widerspruch zwischen Verrechnung (additiv) und Darstellung (multiplikativ) ist nicht sinnvoll bzw. gerechtfertigt. Eine Überprüfung ist in kleinen Datensätzen aus Einzelversuchen bzw. kleinen Serien in der Regel nicht hinreichend möglich, wird aber bei mehrjährigen Auswertungen großer Serien von Sortenversuchen interessant und nützlich.

Eine unberücksichtigte Verletzung der Modellvoraussetzungen führt zu \pm verzerrten Schätzwerten für die Grundgesamtheit, und zwar sowohl bei vollständig als auch bei unvollständig balancierten Daten. Die Problematik steht also grundsätzlich und unabhängig von dem hier vorgestellten Verfahrenskomplex. Allerdings erhöht Unbalanciertheit bei sehr stark variierendem (Ertrags-)Niveau das Verzerrungspotential, wenn die Modellannahmen nicht hinreichend zutreffen sollten.

Zum Verständnis der Problematik kann man folgende Fragen stellen:

Sind z.B. die absoluten Sortendifferenzen [dt/ha] in neuen Versuchen

- a) unabhängig vom Ertragsniveau
- b) proportional zum Ertragsniveau

Für den additiven Ansatz spricht vermeintlich, dass letztendlich Mengen (dt) und nicht Prozente verkauft werden. Es muss aber hinterfragt werden, welche Verallgemeinerungsfähigkeit nachfolgende unterschiedlichen Aussagen für die Praxis mit nicht genau vorhersagbarem Ertragsniveau (neues Jahr und neuer Ort) haben:

- a) Sorte X war im Mittel gegebener Versuche 8 dt/ha ertrags-überlegen
- b) Sorte X war im Mittel gegebener Versuche 10% ertrags-überlegen

Wenn man sich z.B. Pflanzenarten mit geringem, stark schwankendem Ertrag vor Augen führt (z.B. Öllein), dann ist es nicht sehr wahrscheinlich und entspricht auch nicht den Erfahrungen, dass die absoluten Sortenunterschiede auf einem Versuchs-Niveau von z.B. 8 dt/ha gleich hoch ausfallen, wie bei z.B. 30 dt/ha. Wenn hier die Aussage in Richtung b) tendiert, so müssten zumindest relevante multiplikative Anteile vorliegen und es bestünde dann ein Transformationserfordernis.

Es sollte auch erwogen werden, dass evtl. und m.E. mit großer Wahrscheinlichkeit unterschiedliche Varianzkomponenten unterschiedliches Datenverhalten haben, z.B.:

- $\text{var}_{(\text{Sorten})}$ und $\text{var}_{(\text{Sorte} * \text{Umwelten})}$ mit multiplikativen Anteilen
- $\text{var}_{(\text{Rest})}$ additiv, u.U. sogar „über-“ additiv (je höher das Niveau, desto geringer der Fehler)

Wir haben das Ertragsniveau aller Weizen-Sortenversuche in MV langjährig in Beziehung gesetzt a) zum versuchsspezifischen Standardfehler der Einzelwerte und b) zum Variationskoeffizienten (s%). Bei multiplikativem Verhalten hätte s steigen und s% konstant sein müssen. Bei additivem müsste s konstant sein und s% fallen. Die Auswertung ergab signifikanten Abfall von s% mit zunehmendem Ertrag, wobei sogar s tendenziell leicht abfiel (n.s.).

Rechnen mit Absolut- oder Relativwerten (bei stark differierendem Versuchsdurchschnitt):

a) Beispiel und Algorithmus eines additiven Rechengangs mit Absolutwerten

1. Schritt: Bildung des absoluten Serien-Mittelwertes
2. Schritt: Relativierung des Serien-Mittelwertes

Sorte	Einzelversuch 1	Einzelversuch 2	Einzelversuch 3	Serien-MW absolut	Serien-MW relativ
A	67,7	44,3	23,5	45,2	101
B	76,3	39,0	21,6	45,6	102
...
G	72,1	39,0	18,9	43,3	97
MW	72,0	40,7	21,3	44,7	100

b) gleiches Beispiel für den multiplikativen Rechengang mit Relativwerten:

1. Schritt: Relativierung der Einzelversuchs-Mittelwerte aus a)
2. Schritt: Bildung des relativen Serien-Mittelwertes aus den relativen EV-Mittelwerten
(im balancierten Fall identisch mit Auswertung nach der log-Transf. bzw. $\phi=0$)

Sorte	Einzelversuch 1	Einzelversuch 2	Einzelversuch 3	dieser Schritt entfällt hier	Serien-MW relativ
A	94	109	110		104
B	106	96	101		101
...
G	100	96	89		95
100 % = [dt/ha]	72,0	40,7	21,3		44,7

Besonders Sorte 'A' zeigt hier, dass beide Algorithmen auch im balancierten Fall zu deutlich unterschiedlichen Serien-MW kommen.

Der Standpunkt, der additive Ansatz wäre Konvention und daher 'richtig' wird der Frage nicht gerecht: „Wie erreicht man mit dieser Stichproben-Auswertung die bestmögliche Verallgemeinerung für die Grundgesamtheit (regionale Empfehlung für neues Jahr)?“. Die Antwort hängt vom Datenverhalten ab: Ist es additiv ($\phi=1$), so ist a) zu wählen. Ist es multiplikativ ($\phi=0$) ist b) bzw. noch besser die log-Transformation optimal. Bei $\{0<\phi<1\}$ ist eine Zwischenform optimal → umsetzbar durch die unten beschriebene und im Verfahren integrierte Box-Cox-Transformation.

c) Beispiel Gehaltserhöhung zur Veranschaulichung:

Eine Erhöhung um 5 % für alle Lohngruppen ist multiplikativ.

Eine Erhöhung um pauschal 100 € ist additiv.

Es gibt kombinierte Formen, die korrespondiert mit unten beschriebenen Zwischenformen der Transformation zwischen Reinformen der Additivität und Multiplikativität

Nachteile des Rechnens mit Relativzahlen

Bei diesem Vorgehen wird für Relativzahlen ein additives Modell vorausgesetzt, was auf der Ebene der absoluten Daten ein kompliziertes multiplikatives Modell impliziert, wobei die Art des Modells von der Bezugsbasis der Relativzahlen abhängen. Hinzu kommt, dass durch die Berechnung von Relativzahlen stochastische Abhängigkeiten erzeugt werden. Die Auswertung von Relativzahlen ist aus diesen Gründen problematisch.

- setzt multiplikatives Datenverhalten voraus, das in der Regel nicht bzw. nicht in reiner Form gegeben ist
- die Relativwerte aller Sorten reagieren empfindlich auf einzelne auffällige Werte in den ausgewählten Sorten der Bezugsbasis
- je weniger Sorten mehrjährig orthogonal für die Bezugsbasis verfügbar sind, desto empfindlicher reagieren die Relativzahlen
- formales Vorgehen bei Bildung der Bezugsbasis: es müssen u.U. Sorten, für die eigentlich kein Prüfungsbedarf mehr besteht nur zur Sicherung der mehrjährig orthogonalen Bezugsbasis weitergeprüft werden → aufwands-, kosten- und präzisionsrelevant
- langjährige Verrechnung unmöglich, da selbst beim Konzept VRS+VGL mehr als 3-jährige orthogonale Bezugsbasen mit mindestens 3 Sorten selten vorhanden sind
- auch bei einer orthogonalen Bezugsbasis von 3-5 Sorten noch suboptimal gegenüber der Einbeziehung aller verfügbaren Informationen

Alternative, wenn multiplikative Anteile bestehen:

Wenn es Hinweise auf ein (vollständiges oder anteiliges) multiplikatives Zusammenwirken von Effekten gibt, ist die einfachste und beste Reaktion eine Transformation mit anschließender Auswertung nach einem additiven Modell.

→ adäquate Transformation → Verrechnung im additives Modell → Rücktransformation
→ weiterhin erst im letzten Schritt für Darstellungszwecke relativieren

Welche Konsequenzen hätte eine verletzte Additivitäts-Annahme ohne Transformation ?

- bei balancierten Daten:
 - ertraglich unterdurchschnittliche Versuche werden ungerechtfertigt untergewichtet
siehe a); Sorte A
 - ertraglich überdurchschnittlichen Versuche werden ungerechtfertigt übergewichtet
- bei unbalancierten Daten:

die im Zuge der Adjustierung erfolgende „Projektion“ auf das ‚mittlere‘ (Ertrags-)Niveau ist dann gestört, wenn eine Sorte z.B. gehäuft in unter- bzw. überdurchschnittlichen Versuchen stand und gleichzeitig Sorte*Standort(Bodengüte)-Interaktionen auftreten. Die Folge können gewisse Verzerrungen sein (Vergrößerung/Verkleinerung der Effekte), da die „Projektionsstrahlen“ bei verletzter Additivitätsannahme nicht parallel verlaufen; die optimale Transformation stellt Parallelität wieder her und Verhindert die Verzerrung
- bei Einbeziehung von Regionen mit deutlich unterschiedlichem (Ertrags-)Niveau:

z.B. Zielgebiet D-Nord (mittel) mit Nachbargebieten Ostholstein (hoch) und D-Süd (gering)

das unterschiedliche Ertragsniveau der Anbaugebiete könnte ohne adäquate Transformation dazu führen, dass die ertragreichere Region unbegründet über- und die schwache unterbewertet würde

Bei extremen Differenzen im (Ertrags-)Niveau zwischen Umwelten (Einzelversuche, Orte, Jahre, Regionen) ist ohnehin eine fachlich Entscheidung zu treffen, ob einzelne Umwelten zur Grundgesamtheit gezählt werden können oder ob sie als Extreme getrennt ausgewertet und interpretiert werden sollten. In diesem Sinne ist auch der regionale Überlappungsbereich zu begrenzen auf Boden-Klima-Räume, die sich nur graduell aber nicht prinzipiell vom Zielgebiet unterscheiden.

Welche Erfahrungen zum Datenverhalten liegen bei uns bereits vor ?

- Ertragsmerkmale auf hohem Niveau (Winterweizen, Mais konventionell ...)
Schätzwerte im Bereich $\{0,8 \leq \phi \leq 1\}$; unproblematisch
Auswirkungen: gering und nur bei wenigen Sorten mit untypischem Hintergrund und wenigen Ergebnissen (z.B. nach EUSV); immer positive Wirkung und begründbar
- Ertragsmerkmale auf niedrigem Niveau (Leg., S.öfrüchte, ökol. Versuche ...)
Schätzwerte im Bereich $\{0,4 \leq \phi \leq 0,8\}$
etwas deutlichere, gleich bewertete Auswirkungen
- Fungizideffekt (Stufendifferenz) → geringes Niveau
analog zu Ertragsmerkmalen auf niedrigem Niveau
- multiplikativ ermittelte Werte / errechnete Prozentwerte
 - Beispiel: ‚TS-Gehalt‘ bei Silomais
als multiplikativ bestimmtes Merkmal auch so bestätigt $\{0 \leq \phi < 0,5\}$
Auswirkungen: gering (da Differenzierung vgl.weise gering) und nur bei wenigen Sorten; immer positiv und begründbar
 - Beispiel: ‚%lagerndePflanzen‘ bei Mais
noch keine Erfahrung, aber sehr relevante Unterschiede sind aufgrund der starken Differenzierung zu erwarten; Transformation wird nützlich und notwendig sein; darauf weist bei Prozentwerten schon Erna Weber (1986) hin;
die jetzige Praxis der Lager-Auswertungen bei Mais ist ausgesprochen verzerrungsanfällig, da Versuche mit sehr starkem Lager arithmetisch gemittelt werde mit Versuchen mit mittlerem und geringem Befall. Der Mittelwert wird derzeit oft extrem stark von einzelnen Versuchen mit Starklager bestimmt und ist damit sehr fehleranfällig, denn auch hier würde eine ausgewogenere Einbeziehung vieler Umwelten Präzision und Reproduzierbarkeit erhöhen
- sensible Qualitäts-Merkmale,

... die in ihrer Genauigkeit noch deutlich stärker als ‚Ertrag‘ auf suboptimale und heterogene Versuchsbedingungen reagieren und deren Spannweite der Werte relativ gering ist; z.B. ‚Energiekonzentration bei Silomais‘

Ebenso sind z.B. bei Körnermais die Varianzen für TS% bei geringen TS%-Werten (z.B. um 60%) höher, als bei hohen (z.B. um 85%).

{ $\phi \geq 1$ } → wir verwenden $\phi=1$ als absolute Obergrenze

hier scheint es gelegentlich zu einer Überlagerung des Additivitätseffektes durch den ‚Präzisionseffekt‘ zu kommen (s.o.); leider wird im Verfahren ‚PHI‘ die Einzelversuchs-Präzision nicht berücksichtigt, die SE können nicht als Gewichtungsfaktor verwendet werden; die Auswirkung von $\phi > 1$ dürften kaum spürbar sein, da hier i.d.R. das bei diesen Merkmalen das Niveau nominal relativ gering differenziert

Vorschlag: die unterschiedliche Präzision durch Gewichtung nach SE in den Verfahren ‚MW‘ und ‚VK‘ berücksichtigen und hier im Verfahren ‚PHI‘ $\phi=1$ setzen

Es kann verallgemeinert werden, dass eine besondere Relevanz besteht bei:

- geringem (Ertrags-)Niveau
- stark differenziertem (Ertrags-)Niveau zwischen den Versuchen bzw. Regionen
- Prozentwerten u.a. multipl. berechneten Merkmalen

Es soll auch betont werden, dass eine Transformation in den Grenzen $\{0 \leq \phi < 1\}$ unkritisch ist (siehe weiter unten) und sich nur bei einzelnen Sorten nennenswert auswirkt und die Veränderung immer begründbar und positiv zu bewerten war.

Wie kann auf verletzte Additivität reagiert werden ?

Die verwendeten Verfahren arbeiten mit linearen Modellen, setzen also Additivität des Datenverhaltens voraus. Ist dieses nicht gegeben, können Daten so transformiert werden, dass die transformierten Daten additives Verhalten aufweisen. Mit den transformierten Daten können dann die Verfahren optimal arbeiten (z.B. auf der logarithmischen Skala nach einem additiven Modell auswerten). Um im Ergebnis keine abstrakten, dimensionslosen Daten zu erhalten sondern Daten in der üblichen merischen Einheit, sollte eine Rücktransformation sowohl der Mittelwerte als auch der statistischen Parameter erfolgen.

Wie berücksichtigen die Verfahren der ‚Hohenheim-Gülzower‘ Serienauswertung dies ?

Auf Prof. Piephos's Vorschlag wurde die so genannte Box-Cox-Transformation gewählt. Ihr herausragender Vorteil ist die stufenlos anpassbare Transformation.

geschätztes ϕ (ϕ)	verbale Interpretation des Datenverhaltens und Transformation
$\phi \approx 1$	rein additiv → wie untransformierte Daten
$0,5 << \phi << 1$	intermediäres Datenverhalten, näher am additives Verhalten → Box-Cox-Transformation (B-C-T) mit geschätztem ϕ
$\phi \approx 0,5$	intermediäres Datenverhalten → die B-C-T ist identisch mit der Wurzeltransformation
$0 << \phi << 0,5$	intermediäres Datenverhalten, näher am multiplikativen Verhalten → Box-Cox-Transformation (B-C-T) mit geschätztem ϕ
$\phi \approx 0$	rein multiplikativ (entspricht z.B. einer Auswertung über Relativwerte) → die B-C-T ist identisch mit der logarithmischen Transformation
$(\phi > 1)$	wahrscheinlich additiv; liegt 1 im Konfidenzintervall ?, wenn nicht, kann Additivität überlagert sein durch einen „Präzisionseffekt“ (je größer die Merkmalsausprägung, desto kleiner der Fehler) → unser Vorschlag: mit $\phi = 1$ weiterarbeiten und möglichst die Gewichtung nach SE in ‚VK‘ und ‚MW‘ nutzen

Von uns wird i.d.R. aus Gründen der Interpretationsfähigkeit des Datenverhaltens derzeit nur die Arbeit in der Spanne nur $\{0 \leq \phi < 1\}$ empfohlen. In diesen Grenzen ist die Variation von ϕ fachlich als unkritisch anzusehen, da die Schätzwerte dann in jedem Fall innerhalb der beiden plausiblen Extrema liegen: a) erst Relativzahlen bilden und dann mitteln und b) erst Absolutzahlen mitteln und dann Relativieren. Die Zwischenstellung innerhalb der Grenzen $\{0;1\}$ ist als Optimierung im Sinne der „goldenen Mitte“ anzusehen. Selbst wenn das ϕ ein etwas breiteres Konfidenzintervall aufweisen sollte, wird das Risiko minimiert, deutlich falsch zu liegen.

Es handelt sich hier nicht um einen Test auf Einhaltung von Modellvoraussetzungen, sondern um die optimale Anpassung gegebener Daten an die Modelle. Existiert eine signifikante Abweichung von 1, so ergibt sich die Schwierigkeit, die Art der Verletzung näher zu bestimmen (Varianzhomogenität, Normalverteilung, Additivität). Werte von ϕ zwischen 1 und 0 weisen auf einen Übergang zwischen additiver und multiplikativer Wirkung der Effekte hin. Die Trennung der Effekte Additivität, Varianzhomogenität und Normalverteilung war leider nicht möglich. Es ist eher meine Unterstellung, dass Abweichungen ursächlich in Nicht-Additivität begründet sind. Dabei gehe ich davon aus, dass:

$\text{var}_{(\text{Sorte})}$ und $\text{var}_{(\text{S} \cdot \text{Umwelten})}$ einen multiplikativen Anteil enthalten ($\phi < 1$), während sich der Restfehler additiv oder über'-additiv verhält ($\phi \geq 1$; je höher das Ertragsniveau, desto kleiner der mittlere Restfehler).

Als Ausgabe wird der optimale Wert für ‚phi‘ ausgegeben und eine Graphik erstellt. Diese zeigt die Güte des Modells (als Log-Likelihood angegeben) in Abhängigkeit vom Transformationsparameter. Zur Signifikanzprüfung und Ausweisung eines Vertrauensintervalls für den geschätzten Parameter wird eine Linie angezeigt. Alle Transformationsparameterwerte, deren Log-Likelihood größer als die eingezeichnete Linie sind, unterscheiden sich nicht signifikant vom optimalen Transformationsparameter (Differenz der Log-Likelihood ist χ^2 -verteilt mit einem Freiheitsgrad). Interessant ist insbesondere, ob der Wert 1 für den Transformationsparameter (gleich untransformiert) sich vom optimalen Wert unterscheidet.

Erfahrungen aus der „Piepho-Studie“:

In fast allen Fällen waren die Modelle für die transformierten Daten signifikant besser als für die untransformierten Daten.

zur Box-Cox-Transformation:

Eine Möglichkeit, eine Verletzung der Modellannahmen festzustellen und zu berücksichtigen, ist die Transformation. Hier wurde die Klasse der Box-Cox-Transformationen verwendet, die den stufenlos wählbaren Parameter ϕ enthält. Für die Box-Cox-Transformation gilt:

$$y = \begin{cases} \frac{y^\phi - 1}{\phi} & \text{für } \phi \neq 0 \\ \log(y) & \text{für } \phi = 0 \end{cases}$$

wobei ϕ nach der Maximum-Likelihood-Methode geschätzt werden kann. Um die eventuelle Notwendigkeit einer Transformation zu überprüfen, wird die Hypothese getestet, dass die transformierten Daten die Modellannahmen besser erfüllen als die untransformierten Daten. Die jeweilige Güte der Anpassung des Modells an die Daten lässt sich über den Likelihood-Ratio-Test auf der Skala der untransformierten Daten bewerten

$$T_{\text{Vers}} = -2 \cdot (\log L_{\text{red}} - \log L_{\text{erw}}),$$

wobei $T_{\text{Test}} = \chi^2_{x;0,95}$ mit x gleich Anzahl der hinzugefügten Parameter (bei der Box-Cox-Transformation ist $x=1$, also $T_{\text{Test}} = \chi^2_{1;0,95} = 3,84$, da ein Parameter hinzugefügt wird).

Transformationsanleitung für SE-Einzelversuchswerte nach der Delta-Methode:

Werden die Eingangswerte transformiert, so werden automatisch auch die dazugehörigen SE (bei Auswertung mit Gewichtung nach Versuchspräzision) transformiert:

$$w = 1/\text{var}(x) = 1/\text{SE}^2$$

$$1/w1 = \text{var}(y) = \text{var}(x) \cdot [f'(x)]^2 = \text{var}(x) \cdot [x^{2\phi-2}] = \text{SE}^2 \cdot x^{2(\phi-1)}$$

$$\Rightarrow w1 = 1 / \{\text{SE}^2 \cdot x^{2(\phi-1)}\}$$

Rücktransformation:

Um die Versuchsergebnisse und deren Fehler (SE) im Ergebnis in einer metrischen, fachlich interpretierbaren Dimension zu erhalten, werden die Mittelwerte und SE nach dem Modell-Durchlauf automatisch rücktransformiert. Bei den rücktransformierten Werten handelt es sich um Erwartungswerte für den Median, nicht für das arithmetische Mittel. Das ist auch sinnvoll, denn Ursache für die Transformation war ja Nicht-Additivität, also eine schiefe Verteilung. Für so eine Verteilung ist der Median ein besserer Parameter der Lage / des Mittelwertes, als das arithmetische Mittel.

Transformationsanleitungen:

Theorie: Man hat Beobachtungen x , deren Varianz um den Erwartungswert inhomogen ist, also z.B. mit dem Erwartungswert ansteigt (Residuenplot). Die Schätzung einer einheitlichen Varianz ($\text{var}(x)$) ist dann oft unzureichend. Dann kann es sinnvoll sein die Daten zu transformieren. Eine Analyse der mit $f(x)$ transformierten Werte ($y=f(x)$) geben Mittelwertschätzungen und Varianzschätzungen. Zur Interpretation und zur Darstellung kann es notwendig sein, sowohl Mittelwert, als auch Varianz zurück zu transformieren. Für die Mittelwerte ergibt sich die Rücktransformation als Umkehrfunktion von $y=f(x)$, d.h. indem $y=f(x)$ nach x aufgelöst wird. Für die Varianz gilt folgende Beziehung: $\text{var}(x)=1/f'(x)^2 \cdot \text{var}(y)$, d.h. die Varianz eines Mittelwertes hängt vom Niveau der jeweiligen Mittelwerte selber ab.

X = original Y = transformiert

Log-Transformation (z.B. Daten mit Werten an der einseitigen Grenze)

$$y = f(x) = \ln(x)$$

$$f'(x) = \frac{1}{x}$$

$f'(x)$ ist die Ableitung der Funktion $f(x)$. Diese wird benötigt, um die Varianzen zu transformieren (s.o. „Transformationsanleitung für SE-Einzelversuchswerte nach der Delta-Methode“)

Rücktransformation Mittelwert: $x = e^y$

Rücktransformation Varianz: $\text{var}(x) = x^2 \cdot \text{var}(y)$

Logit-Transformation (z.B. Prozentwerte)

$$y = f(x) = \ln\left(\frac{x}{1-x}\right)$$

$$f'(x) = \frac{1}{x(1-x)}$$

Rücktransformation Mittelwert: $x = \frac{1}{e^{-y} + 1}$

Rücktransformation Varianz: $\text{var}(x) = (x - x^2)^2 \cdot \text{var}(y)$

Box-Cox-Transformation ($\lambda \neq 0$)

$$y = f(x) = \frac{x^\lambda - 1}{\lambda}$$

$$f'(x) = x^{\lambda-1} = \frac{x^\lambda}{x}$$

$$\text{Rücktransformation Mittelwert: } x = (\lambda y + 1)^{\frac{1}{\lambda}}$$

$$\text{Rücktransformation Varianz: } \text{var}(x) = \left(\frac{x}{x^\lambda} \right)^2 \cdot \text{var}(y)$$

$$\text{var}(x) = \text{SE}_x^2$$

$$\text{var}(y) = \text{SE}_y^2$$

$$\text{SE}_x = \text{SE}_y \cdot (x / x^{\text{phi}})$$

Beispiel: 2 Behandlungen in 5 Wiederholungen, Daten wurden in SAS mit einem multiplikativen Modell simuliert und nachfolgend leicht modifiziert (Daten siehe Tabelle). Folglich sollte eine log-Transformation verwendet werden.

Behandlung	Wiederholung	x-Wert	f(x)
1	1	9,03	2,20
1	2	9,97	2,30
1	3	5,47	1,70
1	4	7,39	2,00
1	5	6,69	1,90
2	1	29,96	3,40
2	2	22,20	3,10
2	3	27,11	3,30
2	4	18,17	2,90
2	5	20,09	3,00

Die Behandlungsmittelwerte für die transformierten Werte ergeben sich zu 2,02 für Behandlung 1 und 3,14.

$$\frac{2,2 + 2,3 + 1,7 + 2 + 1,9}{5} = 2,02$$

Als Varianz ergibt sich 0,05.

$$\frac{(2,2 - 2,02)^2 + \dots + (3,4 - 3,14)^2 + \dots + (3 - 3,14)^2}{8} = 0,05$$

Mit der oben angegebenen Formel erhält man für die rücktransformierten Mittelwerte 7,54 und 23,10. Diese besitzen eine Varianz von 2,84 (**Beh. 1**) bzw. 26,69 (**Beh. 2**).

$$x = e^{2,02} = 7,54$$

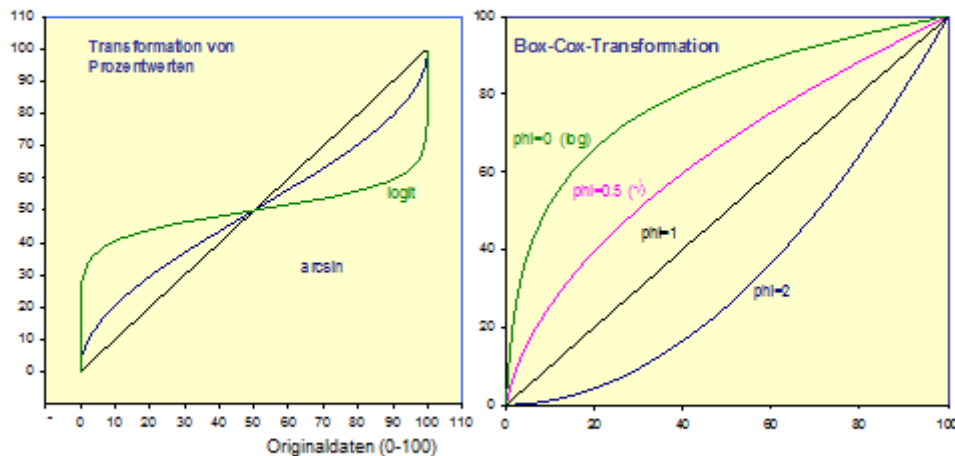
$$\text{var}(x) = 7,54^2 \cdot 0,05 = 2,84$$

Problem: schiefe Verteilung

- Additivität / Normalverteilung wird oft ungeprüft vorausgesetzt, Testung ?
- betrifft eigentlich alle Merkmale, deren *realer* Wertebereich unten und/oder oben an Grenzen stößt
- nur untere Grenze: → häufig log-Transformation
- Prozentwerte:
 - Unterschied zwischen 2% und 6% ist gravierender als 48% gegen 52%
 - Unterschied zwischen 98% und 94% dito
 - **logit- oder arcsin- Transformation möglich, Entscheidung z.B. nach AIC-Kriterium**
- Datenverhalten a-priori unbekannt: → Box-Cox-Transformation

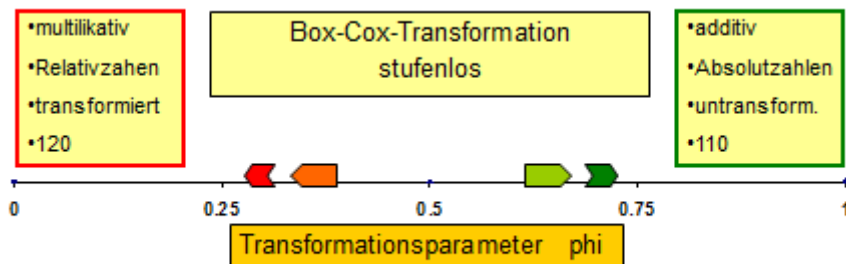
Transformation und Rücktransformation

transformierte Werte (auf 0-100 projiziert)



additives versus multiplikatives Datenverhalten ?

dt/ha	ø Versuche	neuer Anbau
ø VRS	50	100
Sorte X	60	Erwartungswert ohne WW 120 . . . 110



Hanspeter THÖNI,
Institut für Angewandte Mathematik und Statistik, Fachgebiet Biometrie, Universität Hohenheim

Zusammenfassung der Anmerkungen:

Die Modellvoraussetzungen für diese Verfahren sind:

1. Das Modell lässt sich als Linearkombination der Modellparameter schreiben (Lineares Modell).
2. Die Daten sind **homoskedastisch**, d.h. die Varianz der Zufallsvariablen ist unabhängig von deren Erwartungswert.
3. Die Verteilung der Zufallseffekte – im Wesentlichen der Residuen – folgt einer **GAUSS'schen Normalverteilung**.

Die Auswertung erfolgt **immer** unter Annahme eines **linearen** (= additiven) Modells für die **transformierten** Daten.

Wertebereich für ϕ :

In der Abbildung und in der Tabelle wird der Wertebereich unterhalb Null nicht berücksichtigt. Klassische Variablen-Transformationen umfassen auch die reziproke Transformation ($\phi = -1$), welche zur Varianzstabilisierung angezeigt ist, wenn die Varianz proportional zur vierten Potenz des Erwartungswerts ansteigt. Dies trifft häufig zu, wenn z. B. die Messungen sich auf Reaktionsgeschwindigkeiten, Zeitspannen bis zu Erreichen eines bestimmten Endpunktes etc. beziehen.

Der Wertebereich sollte m.E. auf Werte für $\phi < 1$ ausgedehnt werden.

Dagegen ist der Wertebereich $\phi > 1$ nicht so interessant. Mir sind keine Beispiele bekannt, wo eine Abnahme der Varianz mit steigendem Erwartungswert von Relevanz wäre. Kennen Sie Beispiele? (Antwort Michel: Ja; NEL Silomais, Öl Raps, TS% Körnermais → erst, wenn die Werte in den höheren Optimalbereich kommen, werden die Varianzen kleiner). Mir fallen in diesem Zusammenhang nur binomial bzw. beta-verteilte Zufallsvariable ein, d.h. Variable, die auf ein begrenztes Intervall beschränkt sind. Da hilft aber die BOX-COX-Transformation ohnehin nicht weiter.

Das Verfahren von BOX AND COX (1964) beinhaltet weder einen Test auf Linearität des Modells, noch prüft es die Homoskedastizität (**der wichtigsten Voraussetzung für die Anwendung der klassischen Kleinst-Quadrate-Methode**!), noch wird ein Test auf Normalverteilung durchgeführt. Es wird lediglich der Wert der Likelihood unter der Annahme, dass die (transformierten) Daten alle drei Voraussetzungen erfüllen, berechnet, und *unter diesen Voraussetzungen* das Maximum aufgesucht.

Damit erreicht man aber nur, dass nach **dieser** Transformation (d.h. **dieser** Skalenstauchung, falls $\phi < 1$) die Daten durch ein lineares, homoskedastisches Modell am besten beschrieben werden können.

Austausch mit Dr. Schwarzbach (Auszüge):

Von: v.michel@lfa.mvnet.de:

Sehr geehrter Herr Schwarzbach,
Die Ergebnisse bestätigen meine (subjektiven) Eindrücke und Erfahrungen:
1. von einem generellen additiven Datenverhalten ist nicht auszugehen
2. das Datenverhalten der einzelnen Varianzursachen ist unterschiedlich
* Sorteneffekte tendieren zu multiplikativem Verhalten
* Umwelteffekte haben Sie eher als additiv gefunden
(Ihre SRY erinnern mich an das Verfahren der zentrierten Mittelwerte bei Herrn Graf)
* residual ist nicht multiplikativ, sondern fällt mit steigendem Skalenniveau sogar eher ab

Von: Erik Schwarzbach [<mailto:eschwarzbach@iol.cz>]

Sehr geehrter Herr Michel,
schön, dass Sie diese Arbeit interessiert. Sie haben auch das Datenverhalten genau erfasst, Nicht nur Sie, auch die meisten praktischen Züchter gehen intuitiv davon aus, dass das Datenverhalten von Sorten nicht additiv ist, sondern multiplikativ, was in der Beliebtheit von Relativverträgen bei den Züchtern zum Ausdruck kommt. Es lässt sich gut absichern, dass das Datenverhalten von Sorten und Umwelten unterschiedlich ist. Lediglich das Absinken der Residualvarianz mit steigendem Umweltniveau ist wohl ein Zufallseffekt. Die Daten aus Israel zeigen hingegen ein Ansteigen, wenngleich in geringerem Maße als es die genetische Varianz zeigt. Ich halte es für wichtig, dass man das Datenverhalten zunächst empirisch ermittelt und dann erst entsprechende Modelle aufstellt. Viele Theoretiker gingen leider umgekehrt vor, was zu modellbedingten Artefakten führen kann.

Von: v.michel@lfa.mvnet.de:

Sehr geehrter Herr Schwarzbach, was Relativzahlen und Multiplikativität angeht vermute ich einen Fehlschluss a) bei Auswertern und b) bei Anwendern

- a) weil Auswerter in additiven Modellen rechnen,
gehen sie fälschlich a priori auch von additivem Datenverhalten aus
- b) weil Anwender lieber Relativzahlen darstellen,
gehen sie ebenso fälschlich von multiplikativem Verhalten aus

Lieber Volker,

ja, du hast recht, eine solche Standardisierung ist nötig, und das Makro macht diese auch. Eine Standardisierung auf $m=0$, $s=1$ würde es allerdings nicht tun. Was man tun muss, ist zu berücksichtigen, dass man durch die Transformation eine Änderung in der Variable hat, über die bei der Berechnung von Wahrscheinlichkeiten aus der Dichtefunktion zu integrieren ist. Daher ändern sich die Integrationsgrenzen. Die Änderung in der Variablen ist durch die sog. Jacobische Determinante zu berücksichtigen, und genau das macht auch das Makro. Im Anhang schicke ich dir eine Publikation, in deren Anhang das ganz knapp beschrieben ist.
Viele Grüße, Hans-Peter

Literatur

- Atkinson, A.C. 1985. Plots, Transformations and Regression. Oxford University Press, New York
- Gilmour, A.R., B.R. Cullis, S.J. Welham, R. Thomson. 1999. ASREML Reference Manual. NSW Agriculture Biometrie Bulletin No.3. NSW Agriculture, Locked Bag 21, NSW, 2800, Australia, 210ff.
- Piepho, H.-P., V. Michel. 2001. Überlegungen zur regionalen Auswertung von Landessortenversuchen. Informatik, Biometrie und Epidemiologie in Medizin und Biologie 31/4, 123-139

- Michel, V., H.-P. Piepho. 2001. Ertragsauswertung der Sortenversuche in Mecklenburg-Vorpommern. Sommertagung der AG Landwirtschaftliches Versuchswesen der Biometrischen Gesellschaft
- Weber, E. 1986. Grundriß der Biologischen Statistik. VEB Gustav Fischer Verlag. Jena
- Zenk, A., J. Möhring, V. Michel. 2005. Einbindung neuer Methoden zur Routineauswertung von landwirtschaftlichen Versuchen mit Hilfe von SAS-Macros. SAS: Verbindung von Theorie und Praxis, Proceedings der 9. Konferenz der SAS-Anwender in Forschung und Entwicklung. 407-417

Berichte

- Bericht Prüfsystem Winterweizen 24.12.2003. doc
- Prüfsystem Winterweizen 19.03.04. doc
- Prüfsystem Winterweizen 24.08.04.doc
- Prüfsystem WW 24.08.04_Michel.doc