

Bonituren und der Faktor Mensch

Karin Hartung

Institut für Pflanzenbau und Grünland – FG Bioinformatik
Universität Hohenheim

(hartung@uni-hohenheim.de, www.uni-hohenheim.de/bioinformatik/)











18%



50%



18%

Überblick

- Statistische Probleme bei Bonituren
- Menschen und (Prozent-) Bonituren
- Vergleich von 3 Bonituren
 - Accuracy and Precision
 - Differenzen (E-T)
 - Zeit
 - Fragebogen
- Zusammenfassung

Statistische Probleme bei Boniturnoten

Klassengrenzen unbekannt

Klassengrenzen bekannt aber

Klassenmitte nicht unbedingt tatsächlicher Mittelwert der Daten der Klasse

Rücktransformation nicht möglich = Informationsverlust

Änderung der Grenzen über die Jahre

Verlust der Werte der Grenzen

insgesamt: Bonituren schlechter auswertbar als Prozentwerte

Informationsreduktion (überführen in Klassen) nachträglich immer möglich

Menschen und (Prozent-)bonituren

Großer Widerstand

z.B. von Züchterseite „so genau brauchen wir das nicht“ , „das dauert zu lange“
oder von Anfängern „so genau kann ich das nicht“

Richtig machen wollen = Angst vor Fehler

=> das kann ich nicht (gut genug)

=> längeres Überlegen

=> mehr Zeit

„Sinnhaftigkeit“ von Grenzen

log. Unterteilung schwer vorzustellen

leichter : $1/3$, $1/4$, $1/5$ oder 10% und 35%

Stärkere Besetzung von 5% Schritten (Grafik), wenn ungeübt

aber: Wissen um diesen Fehler führt zu Unterbesetzung der 5er

Stimmen die Vorurteile ?

„brauchen wir nicht“
„das dauert zu lange“
„kann ich das nicht“

Was bessert ?

Übung ?
Wissen ?
Vertrauen & Fehlerfreundlichkeit?
???

Vergleich von 3 Bonituren

P1 : Einprozentsschritte (0%, 1%, 2%, 3%,..., 99%, 100%)

P5 : Fünfprozentsschritte (0%, 5%, 10%, ..., 95%, 100%)

R9 : Boniturnoten (1 bis 9, mit hinterliegender log. unterteilter Prozentskala)

| Note | Spanne (%) | | Klassenmitte (%) |
|------|------------|----------|------------------|
| 1 | | 0 | 0 |
| 2 | > | 0 – 2 | 1.0 |
| 3 | > | 2 – 5 | 3.2 |
| 4 | > | 5 – 8 | 6.2 |
| 5 | > | 8 – 14 | 10.6 |
| 6 | > | 14 – 22 | 17.5 |
| 7 | > | 22 – 37 | 28.5 |
| 8 | > | 37 – 61 | 47.5 |
| 9 | > | 61 – 100 | 78.1 |

Annahmen: 1%-Schritte = maximale Genauigkeit

P5 : Mitte = arithmetisches Mittel

R9 : Mitte von 1 & 2 gesetzt Mitte 3-9 = geometrische Mittel der Klasse

Methode

Vorbild: Distrain (Bonitier-Trainingsprogramm für PC von 1988)

- Standardblatt jeweils mit unterschiedlich starkem Befall einer gewählten Krankheit
- eingegebener Schätzwert wird gespeichert für beschreibende Statistiken,
- nach Eingabe erscheint der korrekte Wert

Modellkrankheit: Mehltau

- Simulation durch Ellipsen auf Standardblatt
- Befallsstärke zwischen 1 und 99%
(=> keine Transformationsprobleme mit 0 und 100)
- rechtsschiefe β -Verteilung $\alpha = 1,5$ und $\beta = 3,9$

Datenerhebung (Accessprogramm) :

- Blattanzeige bis Schätzwert eingegeben und bestätigt, dann neues Blatt
- Blätter 1 - 10 zum Kalibrieren mit Angabe des tatsächlichen Befalls
- Blattreihenfolge danach für alle gleich
- im Hintergrund wird die Eingabezeit (Zeit zwischen zwei Bestätigungen) erfasst

Fragebogen :

- Welche Bonituren haben Sie erhoben?
- Welche fiel leichter? Warum?
- Gab es Hemmschwellen, Schwierigkeiten? Wenn wo und welche?
- Welche Bonitur würden sie vorziehen?

Je Person ein oder zwei unterschiedliche Bonituren sowie ein Fragebogen

Erfasst:

Schätzwert

Zeit

subjektiven Faktoren über Fragebogen

Accuracy and Precision

(Genauigkeit und Präzision)

Accuracy: Abweichung des Schätzwertes (E) vom wahren Wert (T)

Precision: Wiederholbarkeit (Campbell and Madden, 1990, p. 110)*.

Statistisch gesehen sind dies Varianz (precision) und Bias (accuracy)

* Campbell C.L., Madden L.V. (1990) Introduction to Plant Disease Epidemiology. Wiley-Interscience, New York.

Lineares Modell

Untersuchte abhängige Variable:

- Zeit logarithmisch transformiert
- Schätzfehlers (D) transformiert:

$$D' = \frac{E - T}{\sqrt{T(1 - T)}}$$

E : Schätzung des Befallswert
T : tatsächlicher Befallswert.

Unterschiedliche Modellwahl für Gruppe A und B

| | Gruppe A | Gruppe B |
|------------|--|---|
| D' | $y_{ijl} = \mu + \alpha_i + \beta_j + \eta_l + \varepsilon x_l + (\alpha\beta)_{ij} + \tau_j x_l + e_{ijl}$ | $y_{ijl} = \mu + \alpha_i + \beta_j + \eta_l + e_{ijl}$ |
| log (Zeit) | $y_{ijl} = \mu + \alpha_i + \beta_j + \eta_l + \varepsilon x_l + (\alpha\beta)_{ij} + \delta_{ij} x_l + e_{ijl}$ | $y_{ijl} = \mu + \alpha_i + \beta_j + \eta_l + \varepsilon x_l + e_{ijl}$ |

y_{ijl} = D' bzw transformierte Zeit

μ = Gesamtmittel, fix

α_i = Effekt der i-ten Methode, fix

β_j = Effekt des j-ten Boniteurs, fix

η_l = Effekt des l-ten Blattes, fix

x_l = wahrer Befallswert des vorhergehenden Blattes [True Previous]

ε = Regressionskoeffizient von True Previous, fix

$(\alpha\beta)_{ij}$ = Interaktion zwischen Methode und Boniteur, zufällig

τ_j = Regressionskoeffizient von Boniteur auf True Previous, fix

δ_{ij} = Regressionskoeffizient von δ_{ij} (Methode und Boniteur) auf True Previous, zufällig

e_{ijl} = Restfehler, mit $e_{ij} \sim N(0, \Sigma)$

Gruppe A : ungeübte Boniteure
Gruppe B : geübte Boniteure

p-Werte der fixen Effekte für transformierte Differenz (Typ I F-Test)

| Effekt | p-Wert of D' | | |
|------------------------------|--------------|----------|--------|
| | Gruppe A | Gruppe B | |
| | | all | 4 Best |
| Blatt | <.0001 | <.0001 | <.0001 |
| Boniteur | 0.2163 | <.0001 | <.0001 |
| Methode | 0.0596 | 0.0028 | 0.0507 |
| True previous | 0.6985 | - | - |
| T.p. x Boniteur ^a | <.0001 | - | - |

^a Interaktion True previous und Boniteur;

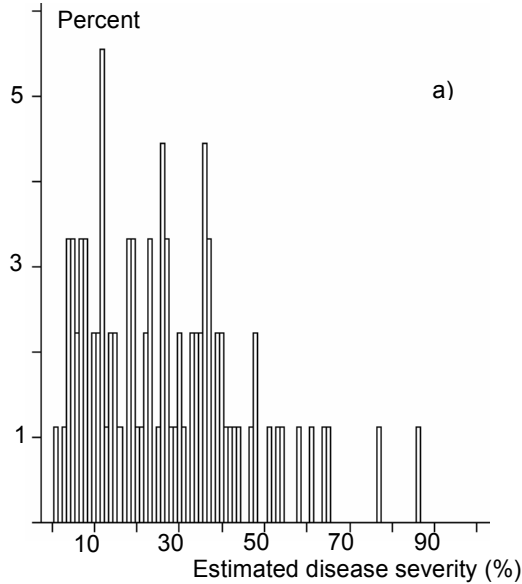
-: nicht Teil des Modells

Histogramm der einzelnen Befallsprozent

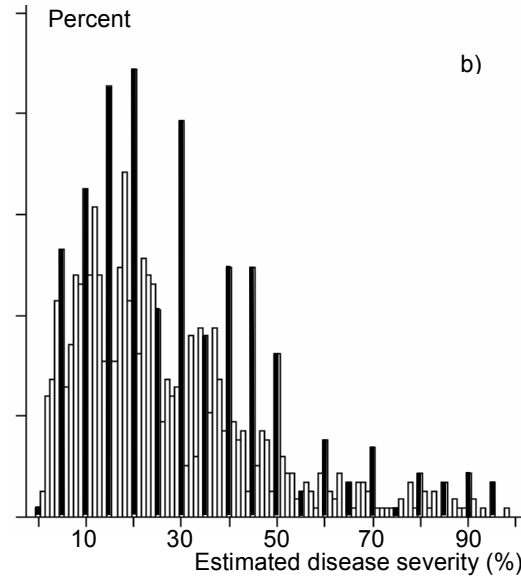
im prozentualen Verhältnis zueinander

(a) rechts-schiefe β -Verteilung der wahren Befallsstärke

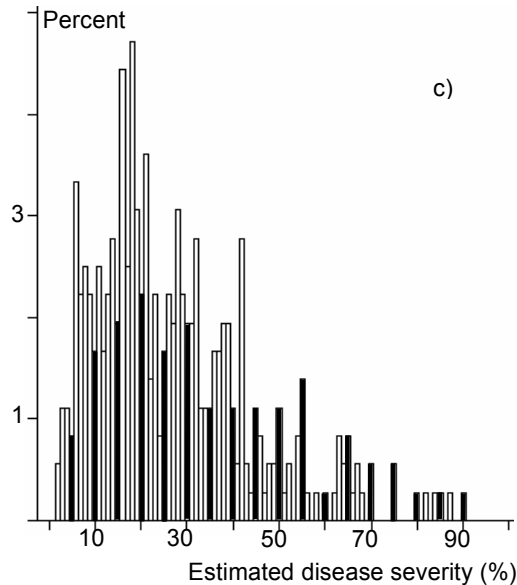
Schwarze Balken zeigen Lage der Vielfache von 5



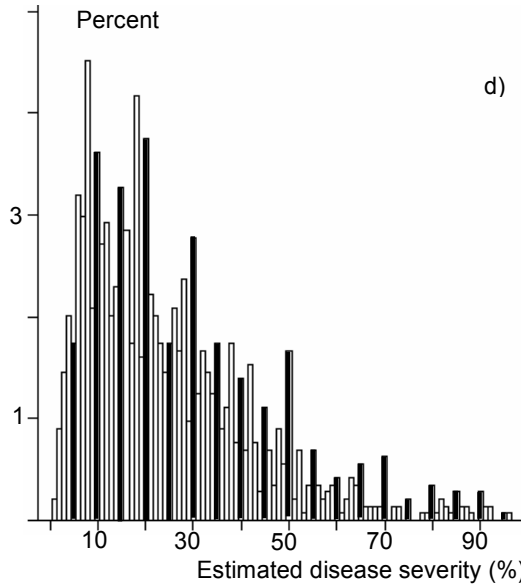
(b) geschätzter Befall Gruppe A (untrainiert)



(c) geschätzter Befall Gruppe B (trainiert)

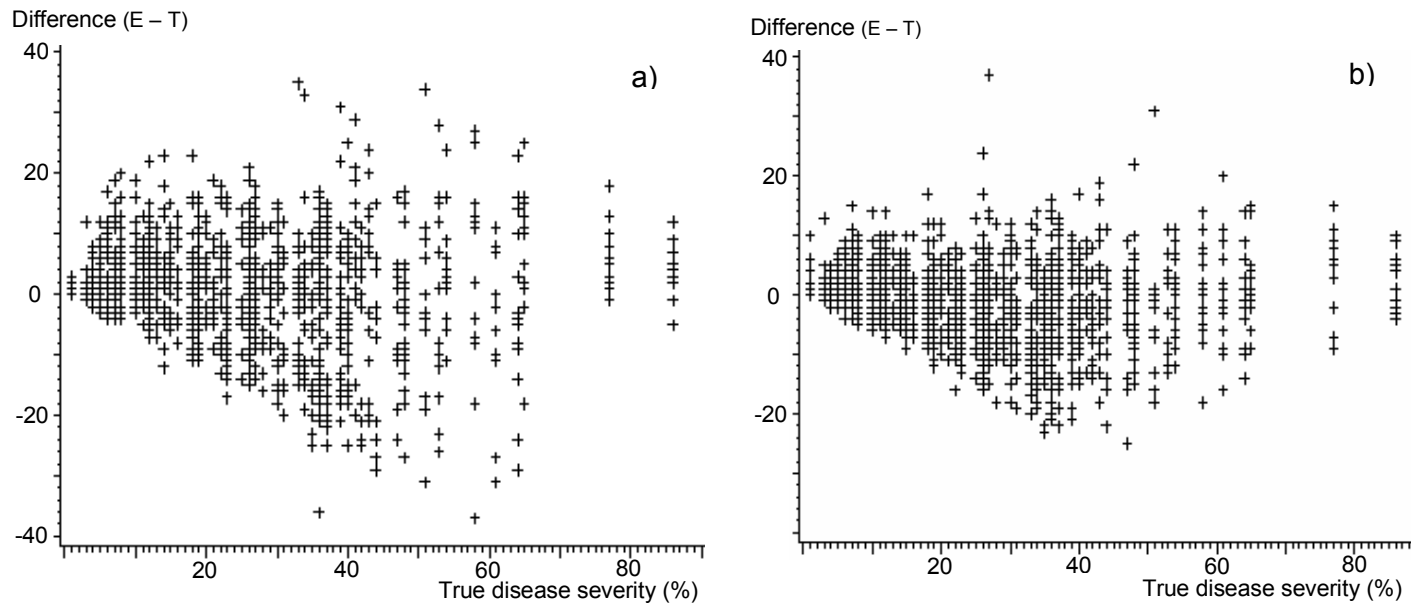


(d) geschätzter Befall der vier Besten



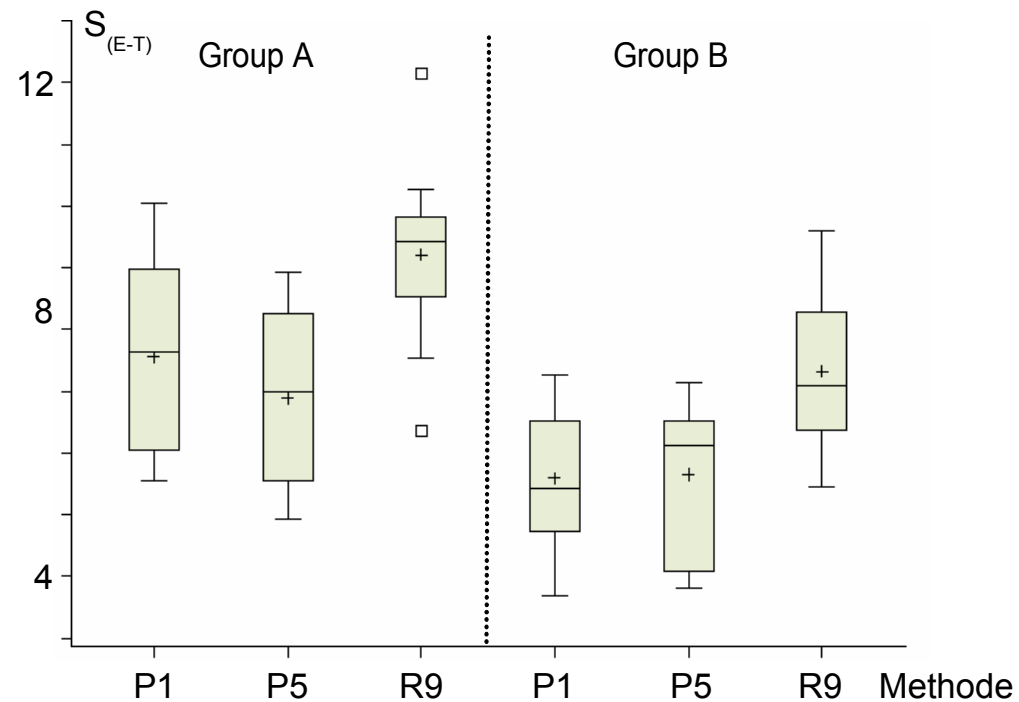
Betrachtung von der Standardabweichung der Differenz (E-T) ($S_{(E-T)} = \text{Accuracy}$)

Typisch elliptische Form der Varianzheterogenität von Prozentbonituren bei beiden Gruppen. Die größte Differenz (T-E) bei mittleren tatsächlichen Werten (T)



Method P1 (a) in Gruppe A (untrainierte), (b) in Gruppe B (trainierte) ein Kreuz kann mehrer Beobachtungen darstellen.

Box-Plots der Standardabweichung der Differenzen (E-T) ($S_{(E-T)}$) der drei Methoden



- Gruppe B besser als Gruppe A (kleinere Mittelwerte)
- Gruppe A Methode R9 zwei stark abweichende Boniteure (kleine Quadrate)

Standardabweichung theoretisch versus real

| Methode | SPS | Gruppe A | | Gruppe B | |
|---------|------|----------------|----------------|----------------|----------------|
| | | $S_{(E-T)min}$ | $S_{(E-T)max}$ | $S_{(E-T)min}$ | $S_{(E-T)max}$ |
| P1 | 0 | 5.54 | 10.06 | 3.68 | 7.25 |
| P5 | 1.45 | 4.94 | 8.93 | 3.81 | 7.14 |
| R9 | 4.80 | 6.36 | 12.14 | 5.46 | 9.59 |

SPS : kleinste mögliche Standardabweichung (theor. Wert bei max. Genauigkeit = 1%)

$S_{(E-T)max}$: Maximum der Standardabweichung der Differenz

$S_{(E-T)min}$: Minimum der Standardabweichung der Differenz

Vergleich der $S_{(E-T)min}$ bzw. $S_{(E-T)max}$ mit SPS :

- Gruppe B immer kleinere Werte als Gruppe A, => Schätzwerte besser
- Differenz $|S_{(E-T)min} - S_{(E-T)max}|$ am kleinsten für P5 (beide Gruppen)

Varianz des Restfehlers (= Precision)

| Method | Varianzschätzer für D' | | |
|--------|------------------------|----------|--------|
| | Gruppe A | Gruppe B | |
| | | all | 4 Best |
| P1 | 0.0260 | 0.0141 | 0.0081 |
| P5 | 0.0231 | 0.0167 | 0.0072 |
| R9 | 0.0355 | 0.0417 | 0.0242 |

- P1 oder P5 am günstigsten

Benötigte Zeit je Bonitur Einfluss der fixen Effekts

| Effekt | p-Wert | |
|---------------|----------|----------|
| | Gruppe A | Gruppe B |
| Methode | 0.0546 | 0.0171 |
| Boniteur | 0.1529 | <.0001 |
| True Previous | 0.5750 | 0.0239 |
| Blatt | <.0001 | <.0001 |

Varianze des Restfehlers (log(sec))²

| Methode | Gruppe A | Gruppe B |
|---------|----------|----------|
| P1 | 0.3313 | 0.3648 |
| P5 | 0.3168 | 0.3601 |
| R9 | 0.2925 | 0.3328 |

Least square (LS) means

| Methode | LS means | |
|---------|-----------------------|-----------------------|
| | Gruppe A ^c | Gruppe B ^c |
| P1 | 1.7472 a | 2.1413 a |
| P5 | 1.5570 ab | 1.7672 b |
| R9 | 1.4054 b | 1.9894 a |

^c : Mittelwerte in einer Spalte mit dem selben Buchstaben sind nicht signifikant verschieden

Fragebogen

| Method | liebste [^] M. | | beste [°] M. | | schlechteste [°] M. | |
|--------|-------------------------|--------|-----------------------|-------|------------------------------|-------|
| | Gr. A | Gr. B* | Gr. A | Gr. B | Gr. A | Gr. B |
| P1 | 3 | 3 | 7 | 6 | 1 | 0 |
| P5 | 6 | 3 | 7 | 4 | 2 | 2 |
| R9 | 5 | 3 | 0 | 0 | 11 | 8 |

[^]: nach Fragebogen

^{*}: ein fehlender Wert

[°]: kleinste bzw. größte Restfehlervarianz je Boniteur

Gruppe A: grobere Methode bevorzugt

genauere Bonitur = bessere Ergebnisse

allerdings eine Person P5 ($S_{(E-T)} = 6.8$) besser als P1 ($S_{(E-T)} = 10.4$).

Gruppe B: Bevorzugte Methode gleich verteilt über alle Skalen.

%-Bonituren waren immer Besser,

allerdings: vier mal P5 besser als P1

zwei mal P5 am schlechtesten aber P1 am besten

Signifikante Unterschiede gibt es nur zwischen R9 und P1 bzw. P5 (p -Wert $<.0001$).

Bemerkenswert: Boniteur geübt mit R9 ($S_{(E-T)} = 8.3$) besser mit P5 ($S_{(E-T)} = 6.3$)

Widerstände und Probleme (nach Fragebogen)

teils gleiche Argumente bei verschiedenen Skalen
widersprachen Aussagen unterschiedlicher Personen zu einem Punkt

- Bei P1 (& P5): genaue Angabe (z.B. 18) ist schwer, besonders zwischen 20 – 80%
- P1: reines Raten, außer wenn Werte <10% und >90%
- leichter wenn alle Blätter gleichzeitig gegeben
- Werte verbesserbar durch zusätzliche Hilfe z.B. Boniturschlüssel
- Größe und Form der Erkrankung beeinflussen den Schätzwert
- P1 einfacher als R9, weil Grenzen R9 ständig nachgeschaut werden müssen
- P5 schwerer als R9
- P5 einfacher auszuwerten als P1, da weniger Klassen
- P5 ist einfach und schnell
- R9 hat komische Klassengrenzen, schwer dran zu gewöhnt => Zeitaufwendig,
- R9 ungenau
- man macht weniger Fehler mit R9, weil Klassen breiter sind

Zusammenfassung

- Vielfache von 5% bei ungeübten verstärkt, Wissen um Problematik => Problemumkehrung; bei Übung nicht mehr auffällig.
- Kleine %-Schritten => Nutzung eines breiteres Wertespektrums, selbst bei Pseudoklassenbildung
- R9 wurde häufig von Boniteuren bevorzugt, besonders von den Untrainierten es richtig machen wollen => weniger Klassen erscheint die Chance größer => Entscheidung fällt leichter => schnell
- „komischen Klassengrenzen“ => verlängern benötigte Zeit bei Geübten, kein sichtbarer Effekt bei Ungeübten
- Kombination verschieden großer Abstände denkbar z.B. 1% für 0-10% u. 90-100%, sonst 5%
- PC-Programme, die beim kalibrieren helfen, sind empfehlenswert

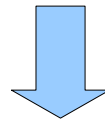
5%-Skala

guter Kompromiss zwischen statistischer Anforderung und Bedürfnissen der Boniteure

Meine Diskussionsfrage

Standardfrage:

Wie kann ich die Daten besser auswerten?



alternative Frage:

Wie kann der Boniteur unterstützt werden, so dass sich die Datenqualität erhöht

Welche Methodik wird in der Medizin genutzt? Ist sie adaptierbar?

Gibt es nützliche Technik?

Wie sieht es mit Sensorik aus?