Analyse von Messfehlern mit zwei Fehlerquellen:

Auswertung mit UNIANOVA in SPSS

H.Hecker

Contents

1 Allgemeine Beschreibung der Fragestellung . . . 1
2 Voraussetzungen und Bezeichnungen . . . 1
3 Schätzung der Varianzkomponeneten . . . 2
3.1 Die Varianz des Ablesefehlers: . . . 2
3.2 Bildfehler und Ablesefehlers . . . 3
3.3 Zusammenfassung . . . 3
4 Auswertung mit SPSS . . . 3
4.1 Fall 1: Daten Objekt-bezogen . . . 3
4.2 Fall 2: Daten Messung-bezogen . . . 3

1. Allgemeine Beschreibung der Fragestellung

Es soll der Messfehler eines Gerätes zur Vermessung von Abständen und Winkeln eines Objektes aus einem Röntgenbild bestimmt werden. Dabei wird davon ausgegangen, daß der gesamte Messfehler sich aus 2 Komponenten zusammensetzt:

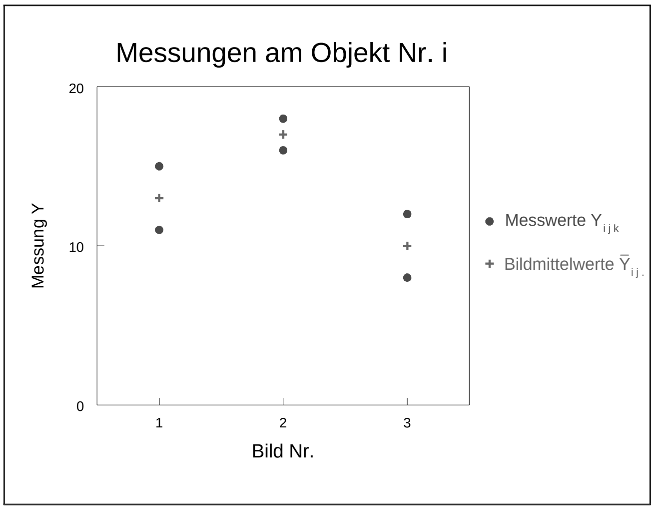
1. **Der Bildfehler**: Die Aufnahme des Bildes selber erzeugt zufallsabhängige Schwankungen: bei identischem Objekt (z.B.: Oberkiefer eines Patienten) unterscheiden sich zwei unabhängig durchgeführte Aufnahmen voneinander.
2. **Der Ablesefehler**: Bei identischem Bild unterscheiden sich zwei unabhängig (z.B. von verschiedenen Personen) durchgeführte Ablesungen.

Die Aufgabe besteht darin, aus vorliegenden Daten den Gesamtfehler sowie die Größe der einzelnen Fehlerkomponenten zu schätzen.

2. Voraussetzungen und Bezeichnungen

- Es wird angenommen, daß die Messungen an n verschiedenen Objekten durchgeführt werden.
- Zu jedem Objekt werden J Bilder erzeugt.
- Zu jedem Bild werden K Messungen durchgeführt.

Es bezeichnet dann Y_{ijk} das Ergebnis der Messung Nr. k zum Bild Nr. j am Objekt Nr. i.:



Aufgrund der Unabhängigkeit aller Bild-Aufnahmen und Ablesevorgängen werden dann folgende Annahmen gemacht:

$$
\begin{aligned}
Y_{ijk} &= A_i + B_{ij} + C_{ijk} \quad (i = 1, \dots, n; \quad j = 1, \dots, J; \quad k = 1, , \dots, K) \quad \text{wobei} \\
A_i &= \text{"Wahrer Wert" von Objekt Nr. } i, \\
B_{ij} &= \text{Bildfehler von Bild Nr. } j \text{ am Objekt Nr. } i, \\
C_{ijk} &= \text{Fehler der Ablesung Nr. } k \text{ am Bild Nr. } j \text{ bei Objekt Nr. } i\text{, mit} \\
E(B_{ij}) &= E(C_{ijk}) = 0\ , \\
var(B_{ij}) &= \sigma_B^2 \text{ und} \\
var(C_{ijk}) &= \sigma_C^2 \text{ und:} \\
& \quad \text{alle } B_{ij}, C_{ijk} \text{ stochastisch unabhängig}
\end{aligned}
$$

Gesucht ist die Gesamt–Varianz der Fehler :

$$\sigma^2 = \sigma_B^2 + \sigma_C^2$$

sowie deren Komponenten σ_B^2 und σ_C^2.
Weitere Bezeichnungen:

$$\begin{aligned}
\overline{Y}_{ij.} &= \frac{1}{K}\sum_k Y_{ijk} = \text{Mittelwert zum Bild } j \text{ von Objekt } i, \\
&= \frac{1}{K}(\sum_k A_i + \sum_k B_{ij} + \sum_k C_{ijk}) \\
&= A_i + B_{ij} + \overline{C}_{ij.} \quad \text{mit} \\
\overline{C}_{ij.} &= \frac{1}{K}\sum_k C_{ijk} \\
\overline{Y}_{i..} &= \frac{1}{JK}\sum_{j,k} Y_{ijk} = \frac{1}{J}\sum_{j,k}\overline{Y}_{ij.} = \text{Mittelwert Objekt } i \\
&= \frac{1}{JK}(\sum_{j,k} A_i + \sum_{j,k} B_{ij} + \sum_{j,k} C_{ijk}) \\
&= A_i + \overline{B}_{i.} + \overline{C}_{i..} \quad \text{mit} \\
\overline{B}_{i.} &= \frac{1}{J}\sum_j B_{ij} \quad \text{und} \\
\overline{C}_{i..} &= \frac{1}{JK}\sum_{j,k} C_{ijk} = \frac{1}{J}\sum_{j,k}\overline{C}_{ij.} \qquad (2.1)
\end{aligned}$$

3. Schätzung der Varianzkomponeneten

3.1. Die Varianz des Ablesefehlers:

Es ist für jedes i und j :

$$\begin{aligned}
Y_{ijk} - \overline{Y}_{ij.} &= (A_i + B_{ij} + C_{ijk}) - (A_i + B_{ij} + \overline{C}_{ij.}) \\
&= C_{ijk} - \overline{C}_{ij.}
\end{aligned}$$

und daher

$$\begin{aligned}
E\left[\frac{1}{K-1}\sum_k (Y_{ijk} - \overline{Y}_{ij.})^2\right] &= var(C_{ijk}) \\
&= \sigma_C^2 \quad (i = 1, \ldots, n; \quad j = 1, \ldots, J)
\end{aligned}$$

Für σ_C^2 wählt man daher insgesamt die folgende erwartungstreue Schätzung:

$$\begin{aligned}
\hat{\sigma}_C^2 &= \frac{1}{nJ}\sum_{i,j}\left[\frac{1}{K-1}\sum_k (Y_{ijk} - \overline{Y}_{ij.})^2\right] \qquad (3.1) \\
&= \frac{1}{nJ(K-1)}\sum_k (Y_{ijk} - \overline{Y}_{ij.})^2
\end{aligned}$$

3.2. Bildfehler und Ablesefehlers

Man betrachtet die Varianz der Bild-Mittelwerte $\overline{Y}_{ij.}$:

$$\begin{aligned}\overline{Y}_{ij.} - \overline{Y}_{i..} &= (A_i + B_{ij} + \overline{C}_{ij.}) - (\ A_i + \overline{B}_{i.} + \overline{C}_{i..} \quad) \\ &= (B_{ij} - \overline{B}_{i.}) + (\overline{C}_{ij.} - \overline{C}_{i..})\end{aligned}$$

Wegen der Unabhängigkeit der B_{ij} und C_{ijk} ist dann

$$\begin{aligned}E\left[\frac{1}{J-1}\sum_j(\overline{Y}_{ij.} - \overline{Y}_{i..})^2\right] &= E\left[\frac{1}{J-1}\sum_j(\ B_{ij} - \overline{B}_{i.})^2\right] + \frac{1}{J-1}\left[\sum_j(\ \overline{C}_{ij.} - \overline{C}_{i..}\ )^2\right] \\ &= var(B) + var(\ \overline{C}_{ij.}) \\ &= \sigma_B^2 + \frac{1}{K}var(C) \\ &= \sigma_B^2 + \frac{1}{K}\ \sigma_C^2\end{aligned}$$

Insgesamt folgt (bei Mittelung über alle n Objekte:)

$$E\left[\frac{1}{n(J-1)}\sum_{i,k}(\overline{Y}_{ij.} - \overline{Y}_{i..})^2\right] = \sigma_B^2 + \frac{1}{K}\ \sigma_C^2 \tag{3.2}$$

3.3. Zusammenfassung

Erwartungstreue Schätzungen erhält man nach (3.1) und (3.2) wie folgt:

Parameter	Schätzung	Kürzel
σ_C^2	$\frac{1}{nJ(K-1)}\sum_k(Y_{ijk} - \overline{Y}_{ij.})^2$	Q_1
$\sigma_B^2 + \frac{1}{K}\ \sigma_C^2$	$\frac{1}{n(J-1)}\sum_{i,k}(\overline{Y}_{ij.} - \overline{Y}_{i..})^2$	Q_2
σ_B^2	$Q_2 - \frac{1}{K}Q_1$	
$\sigma^2 = \sigma_B^2 + \sigma_C^2$	$Q_2 + (1 - \frac{1}{K})Q_1$	

4. Auswertung mit SPSS

4.1. Fall 1: Daten Objekt-bezogen

Die Daten enthalten für jedes Objekt eine Zeile mit den $J \times K$ Variablen V_{jk} ($V_{jk} \hat{=} Y_{ijk}$ in Objekt i). Dann bildet man per COMPUTE-Anweisung die Mittelwerte der Messungen je Bild: $\overline{V}_{j.}$, ebenso die Mittelwerte je Objekt: $\overline{V}_{..}$. Daraus sind dann die Residuen der Ablesungen je Bild: $V_{jk} - \overline{V}_{j.}$, und die Residuen der Objekt-Mittelwerte: $\overline{V}_{j.} - \overline{V}_{..}$, ebenso deren Quadrate, zu bilden. Die Häufigkeitsauszählungen liefern dann die Summen über alle Objekte, die nach der obigen Tabelle noch "geeignet" zu dividieren sind (durch $nJ(K-1)$ bzw. $n(J-1)$). Somit erhält man Q_1 und Q_2 und damit nach der Tabelle die gesuchten Parameter-Schätzungen.

4.2. Fall 2: Daten Messung-bezogen

Die Daten enthalten für jede Messung eine Zeile, in der zur weiteren Kennzeichnung der Messung:

- das Objekt (i : Variable "`objekt`")
- die Nummer des Bildes (j: Variable "`bild`"), und
- die Nummer der Messwiederholung innerhalb jedes Bildes (k)

angegeben ist. Man erhält dann die benötigten Quadratsummen aus der Varianzanalyse, wenn man das folgende Modell ansetzt:

```
UNIANOVA
    Y by objekt bild
    /METHOD = SSTYPE(3)
    /INTERCEPT = INCLUDE
    /CRITERIA = ALPHA(.05)
    /DESIGN = objekt objekt*bild .
```

Beispiel mit $n = 16$, $J = 2$ und $K = 2$:

Tests der Zwischensubjekteffekte

Abhängige Variable: Y

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz
Korrigiertes Modell	1477.187[a]	31	47.651	73.116	.000
Konstanter Term	145685.348	1	145685.348	223540.212	.000
OBJEKT	1440.330	15	96.022	147.337	.000
OBJEKT * BILD	36.857	16	2.304	3.535	.001
Fehler	20.855	32	.652		
Gesamt	147183.390	64			
Korrigierte Gesamtvariation	1498.042	63			

a. R-Quadrat = .986 (korrigiertes R-Quadrat = .973)

Bei der Umsetzung ist zu beachten, wie die mittleren Quadrate gebildet sind:

Quelle	Mittel der Quadrate	entspricht:
Fehler	$\frac{1}{nJ(K-1)} \sum_{i,j,k}(Y_{ijk} - \overline{Y}_{ij.})^2$	Q_1
Objekt × Bild	$\frac{1}{n(J-1)} K \sum_{i,j}(\overline{Y}_{ij.} - \overline{Y}_{i..})^2$	KQ_2

Daraus ergeben sich die folgenden Schätzungen:

Parameter	Schätzung	Kürzel
σ_C^2	Mittel der Quadrate "Fehler"	Q_1
$\sigma_B^2 + \frac{1}{K}\,\sigma_C^2$	$\frac{1}{K}\times$ Mittel der Quadrate "Objekt × Bild"	Q_2
σ_B^2	$Q_2 - \frac{1}{K}Q_1$	
$\sigma^2 = \sigma_B^2 + \sigma_C^2$	$Q_2 + (1 - \frac{1}{K})Q_1$	

Im obigen Beispiel lauten die Parameterschätzungen also:

$$
\begin{aligned}
Q_1 &= 0.652 \\
Q_2 &= \frac{1}{2} \times 2.304 \\
\widehat{\sigma}_C^2 &= 0.652 \\
\widehat{\sigma}_B^2 &= \frac{1}{2} \times 2.304 - \frac{1}{2} \times 0.652 = .826 \\
\widehat{\sigma}^2 &= \widehat{\sigma}_B^2 + \widehat{\sigma}_C^2 = .826 + 0.652 = 1.478 \\
&= \frac{1}{2} \times 2.304 + (1 - \frac{1}{2})0.652
\end{aligned}
$$