

# Kapitel 3

## Zuchtwertschätzung

### 3.1 Einführung

Die Züchtung von Nutztieren basiert auf der künstlichen **Selektion** und auf der gezielten **Anpaarung** von Elterntieren zur Erzeugung der nachfolgenden Generation. Selektion ist nur möglich, wenn ein Kriterium existiert, wonach die zu selektierenden Tiere rangiert werden können.

Nehmen wir, dass wir für eine Gruppe von 10 Tiere je einen bestimmten Wert eines gemessenen Kriteriums beobachten. Dann sieht ein Plot (siehe Abbildung 3.1) davon wie folgt aus.

Wenn wir die Werte des in Abbildung 3.1 gezeigten Plots als Selektionskriterium verwenden wollen, dann müssen wir als erstes die Tiere gemäss den Werten des Kriteriums ordnen. Das Resultat dieser Rangierung sieht dann wie folgt aus

#### 3.1.1 Geschichte

Zuerst wurden Tiere nach ihrer äusseren Erscheinung, d.h. also nach ihrem **Phänotyp** ausgewählt. Für Typus- und Erscheinungsmerkmale funktionierte das gut, und dies führte zu einer Differenzierung in sehr viele verschiedene Nutztierassen. Der Grund dafür ist, dass diese Merkmale oft nur monogen oder durch nur wenige Gene beeinflusst werden.

Bei den Leistungsmerkmalen war diese Strategie weniger erfolgreich, da die Leistungen in der Vergangenheit nicht objektiv gemessen wurden, sondern nur geschätzt werden konnten. Erst mit der Einführung der

Tabelle 3.1: Rangierung der Tiere nach deren Werte des Kriteriums

Tier	Kriterium
Tier 5	92
Tier 10	94
Tier 4	98
Tier 7	98
Tier 8	99
Tier 2	101
Tier 3	103
Tier 6	106
Tier 1	110
Tier 9	113

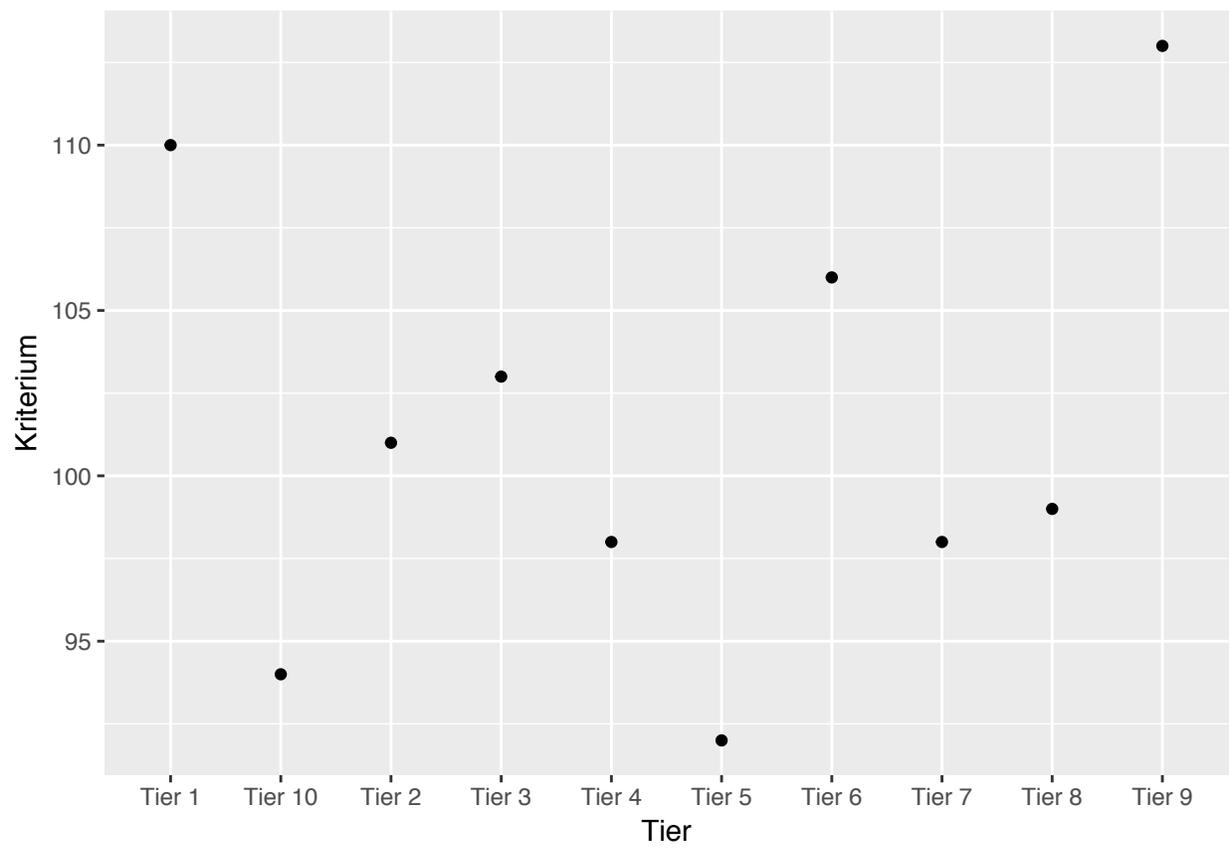


Abbildung 3.1: Ungeordnete Werte eines Kriteriums für 10 Tiere

Tabelle 3.2: Zahlenbeispiel für die Zuchtwertschätzung (tägliche Zunahmen bei Fleischrindern in kg/Tag)

Tier	Vater	Mutter	Betrieb	Zunahme
12	1	4	1	1.26
13	1	4	1	1.32
14	1	5	1	1.40
15	1	5	1	1.44
16	1	6	2	1.52
17	1	6	2	1.50
18	1	7	2	1.42
19	1	7	2	1.46
20	2	8	1	1.34
21	2	8	1	1.32
22	2	9	1	1.24
23	2	9	1	1.28
24	2	10	2	1.44
25	2	10	2	1.40
26	3	11	2	1.54
27	3	11	2	1.56

Tabelle 3.3: Mittelwert der Zunahmen der Nachkommen pro Vater

Vater	Mittelwert
1	1.42
2	1.34
3	1.55

**Leistungsprüfung** und den Auswertungen anhand von modernen Zuchtwertschätzverfahren in der zweiten Hälfte des 20. Jahrhunderts konnte eine grosse Leistungssteigerung beobachtet werden. Dabei gibt es auch zu beachten, dass die grosse Leistungssteigerung auch mit der Verfügbarkeit von billigen **Rechenressourcen** zusammenfällt. Die modernen Zuchtwertschätzverfahren verlangen das effiziente Lösen von sehr grossen Gleichungssystemen und das kann nur mit Hilfe des Computers gemacht werden.

### 3.1.2 Zahlenbeispiel

Die verschiedenen Zuchtwertschätzverfahren sollen anhand des Zahlenbeispiels in Tabelle 3.2 aus dem Bereich der Fleischrinderzucht dargestellt werden.

Werden die Väter nach den phänotypischen Mittelwerten ihrer Nachkommen geordnet lautet die Reihenfolge 3, 1, 2. Die konkreten Mittelwerte sind in Tabelle 3.3 gezeigt.

Bei der Rangierung aufgrund der phänotypischen Mittelwerte besteht eine gewisse Unsicherheit, da Vater 3 nur Nachkommen auf Betrieb 2 hat. Da das Management mit Fütterung und Haltung auf den Betrieben unterschiedlich ist, kann das zu Unterschieden in den phänotypischen Leistungen führen, welche nichts mit dem genetischen Potential der Väter zu tun haben. In unserem Zahlenbeispiel stellt der Betrieb ein sogenannter systematischer Umwelteinfluss oder -faktor dar. Damit wir unsere Rangierung der Selektionskandidaten (in unserem Beispiel sind das beispielsweise die Väter) rein auf ihrem genetischen Potential machen können, müssen wir einen Weg finden, um systematische Umwelteinflüsse zu korrigieren.

Die Tiere haben auch eine unterschiedliche Anzahl an Informationen für die Zuchtwertschätzung verfügbar.

- Die Väter haben 8, 6, bzw. 2 Nachkommen, welche Informationen liefern.
- Jede Mutter hat 2 Nachkommen.
- Die Nachkommen haben jeweils ein Vollgeschwister und zwischen 0 bis 6 Halbgeschwister.

Im Folgenden werden verschiedene Zuchtwertschätzverfahren für die Väter und dann für die Mütter vorgestellt. Für die weiteren Berechnungen brauchen wir noch ein paar Parameter.

- Für die phänotypischen Varianz nehmen wir einfach die aus den Daten berechnete empirische Varianz. Diese beträgt  $\sigma_p^2 = 0.0102(kg/Tag)^2$
- als Heritabilität nehmen wir einen Wert von  $h^2 = 0.25$  an
- Die genetisch-additive Varianz berechnet sich als  $\sigma_a^2 = h^2 * \sigma_p^2 = 0.25 * 0.0102 = 0.0026(kg/Tag)^2$

### 3.1.3 Definition des Zuchtwerts

Als erstes wollen wir den Begriff des **Zuchtwerts** definieren. Aus den mehreren Definitionen verwenden wir diese, welche es uns erlaubt ein Schätzverfahren daraus abzuleiten. Eine anschauliche und praktisch Relevante Definition lautet

**Definition 3.1** (Zuchtwert). Der Zuchtwert eines Tieres  $s$  entspricht der doppelten Leistungsabweichung der Nachkommen<sup>1</sup> des Tieres  $s$  im Vergleich mit dem Populationsdurchschnitt.

Aus Definition 3.1 ergeben sich ein paar spannende Fakten. Diese sind im folgenden Abschnitt kurz zusammengefasst.

- In der Zuchtwertdefinition wird die doppelte Leistungsabweichung verwendet, da jedes Tier nur die Hälfte seiner Allele an die Nachkommen weitergibt.
- Die für den Zuchtwert verwendete Leistungsabweichung muss auf der Basis von *vielen Nachkommen* erhoben werden. Es bekommt zwar jeder Nachkomme die Hälfte der Allele von den Eltern, aber es bekommt nicht jeder Nachkomme die gleichen Allele von den Eltern. Die Allele in einem Nachkommen stellen eine *Zufallsstichprobe* der Elternallele dar. Deshalb braucht es viele Nachkommen, damit der Zuchtwert sich stabilisiert. Deshalb können weder Zuchtwerte noch Leistungen einzelner Nachkommen vorausgesagt werden.
- Da der Zuchtwert als Abweichung von einem Populationsdurchschnitt definiert ist, gilt er auch nur in dieser Population. Zuchtwerte per-se lassen sich also nicht über Populationen hinweg vergleichen. Es gibt spezielle Verfahren, wie Zuchtwerte durch eine Umrechnung vergleichbar gemacht werden können.
- Aus der Definition des Zuchtwerts als Abweichung von einem Durchschnitt, muss der Durchschnitt der Zuchtwerte in einer Population wieder bei 0 sein.

## 3.2 Prinzip der Zuchtwertschätzung

In der Zuchtwertschätzung werden beobachtete Leistungen mit zwei Operationen verarbeitet.

1. Die Leistungen werden **relativiert**, d.h. sie werden mit anderen Leistungen, welche unter ähnlichen Bedingungen erbracht wurden, miteinander verglichen. Rechnerisch passiert das Relativieren durch die Subtraktion eines Vergleichswertes von den Leistungen.
2. Die Leistungen werden **gewichtet**, da für verschiedene Tiere unterschiedlich viele Informationen zur Verfügung stehen. Intuitiv ist klar, dass wir einem Zuchtwert aufgrund von einer grossen Anzahl von Nachkommen mehr vertrauen, als einem der nur aufgrund von sehr wenigen Nachkommen geschätzt wurde. Mathematisch entspricht die Gewichtung einer Multiplikation mit einem Faktor.

Fassen wir diese beiden Operationen in einer Formel zusammen so erhalten wir

$$\hat{u} = b * (y - \mu) \tag{3.1}$$

---

<sup>1</sup>Diese Definition führt erst dann zu einem konsistenten Zuchtwert, wenn es sich hier um eine grosse Anzahl von Nachkommen handelt.

wobei

- $\hat{u}$  Zuchtwert
- $b$  Gewichtungsfaktor
- $y$  Leistung
- $\mu$  Vergleichswert

Das hier vorgestellte Prinzip gilt für alle konventionellen Zuchtwertschätzungen. Sie unterscheiden sich nur, wie die Faktoren  $b$  und die Vergleichswerte  $\mu$  bestimmt werden.

### 3.3 Geschätzte Zuchtwerte aufgrund einer Eigenleistung

Die Zuchtwertschätzung (ZWS) aufgrund einer Eigenleistung ist der einfachste Fall. Das Prinzip der ZWS lässt sich an diesem Fall sehr gut demonstrieren. Das Ziel ist, dass wir für die 16 Nachkommen in Tabelle 3.2 einen Zuchtwert aufgrund der Eigenleistung im Merkmal Zunahme schätzen.

#### 3.3.1 Schritt 1: Relativieren

Im ersten Schritt relativieren wir die Leistungen, damit wir diese zwischen den beiden Betrieben vergleichen können. Als Vergleichswert  $\mu$  der Betriebe verwenden wir deren Least Squares Lösungen. Deren Berechnung ist im Abschnitt 3.5 gezeigt. Für das Relativieren berechnen wir

$$e = y - X\beta \tag{3.2}$$

wobei

- $y$  Vektor mit Leistungen
- $\beta$  Vektor mit Least Squares Lösungen der Betriebe
- $X$  Design-Matrix, welche Betrieb mit Beobachtungen verknüpft

Der Vektor  $e$  entspricht eigentlich den Residuen aus dem Modell in (3.24) und sind in der folgenden Tabelle gezeigt

Tier	Residuen
12	-0.065
13	-0.005
14	0.075
15	0.115
16	0.040
17	0.020
18	-0.060
19	-0.020
20	0.015
21	-0.005
22	-0.085
23	-0.045
24	-0.040
25	-0.080
26	0.060
27	0.080

### 3.3.2 Schritt 2: Gewichtung der relativen Leistungen

Für die Gewichtung der relativierten Leistungen müssen wir wissen, welcher Teil der phänotypischen Überlegenheit genetisch bedingt ist. Hierzu benötigen wir die Regression des Genotyps ( $u$ ) auf den Phänotyp  $y$  eines Individuums. Der Regressionskoeffizient  $b_{u,y}$  des Genotyps auf den Phänotyp entspricht der Erblichkeit  $h^2$ . Es gilt also

$$b_{u,y} = \frac{\sigma_{u,y}}{\sigma_y^2} = h^2 \quad (3.3)$$

wobei

$b_{u,y}$	Regressionskoeffizient des Genotyps auf den Phänotyp
$\sigma_{u,y}$	Covarianz zwischen Genotyp und Phänotyp
$\sigma_y^2$	Phänotypische Varianz

Als Resultat erhalten wir die geschätzten Zuchtwerte aufgrund von Eigenleistungen als

$$\hat{u} = h^2 * e = h^2 * (y - X\beta) \quad (3.4)$$

Für unser Zahlenbeispiel haben wir dann

Tier	Zuchtwert
12	-0.01625
13	-0.00125
14	0.01875
15	0.02875
16	0.01000
17	0.00500
18	-0.01500
19	-0.00500
20	0.00375
21	-0.00125
22	-0.02125
23	-0.01125
24	-0.01000
25	-0.02000
26	0.01500
27	0.02000

Die Rangierung der Tiere aufgrund der Residuen und aufgrund der geschätzten Zuchtwerte ist die gleiche. Die Multiplikation mit einem konstanten Faktor ändert die Rangierung nicht. Die Streuung der geschätzten Zuchtwerte  $\hat{u}$  ist aber geringer als jene der Residuen. Der Grund dafür ist, dass der Faktor ( $h^2$ ) kleiner ist als 1.

Die Verwendung dieses einfachen Zuchtwertschätzverfahrens ist sehr eingeschränkt. Es bekommen nur Tiere mit einer Eigenleistung einen Zuchtwert. Die Tiere, wie die Väter oder die Mütter in unserem Zahlenbeispiel bekommen keinen geschätzten Zuchtwert, da sie keine Eigenleistung haben. Unser Ziel für die Entwicklung von anderen Verfahren soll sein, dass alle Tiere einen Zuchtwert bekommen und dass dafür alle verfügbare Information verwendet werden soll.

## 3.4 Selektionsindex zur Schätzung von Zuchtwerten

Es bestehen zwei wesentliche Methoden für die Schätzung von Zuchtwerten.

1. **Selektionsindextheorie** nach (Hazel and Lush, 1943) und (Hazel, 1943) und
2. das wesentlich aufwändigere **BLUP**-Verfahren nach (Henderson, 1973)

Beide Verfahren basieren auf der gleichen genetischen Theorie. Sie unterscheiden sich aber in der Art, wie sie die systematischen Umwelteffekte korrigieren.

### 3.4.1 Einführung zur Selektionsindextheorie

Grundsätzlich besteht die Zuchtwertschätzung darin, für einen Selektionskandidaten den auf additiver Genwirkung beruhenden Anteil der Leistungsüberlegenheit bei sich und den verwandten Tieren zu schätzen und so miteinander zu kombinieren, dass die Korrelation zwischen dem geschätzten und dem wahren Zuchtwert maximiert werden. Alternativ zur Maximierung der Korrelation kann auch die Varianz des Fehlers zwischen geschätzten und wahren Zuchtwert minimiert werden. Aufgrund der soeben formulierten Vorgaben wird klar, dass wir immer zwischen zwei verschiedenen Zuchtwerten unterscheiden müssen.

1. Der **wahre Zuchtwert** eines Individuums, welcher der Summe der Gensubstitutionseffekte des Selektionskandidaten entspricht und
2. Der **geschätzte Zuchtwert**, der mit statistischen Methoden aus verschiedenen Informationsquellen geschätzt wird und somit immer fehlerbehaftet ist. Die Größenordnung des Fehlers lässt sich abschätzen, aber der Schätzfehler kann nicht vermieden werden.

Der wahre Zuchtwert ist eine Funktion der durchschnittlichen Einzelgensubstitutionseffekten und ist somit von den Allelfrequenzen in einer Population abhängig. Der wahre Zuchtwert eines Tieres ist also keine konstante Grösse, sondern verändert sich mit den Änderungen einer Population. Dies ist mit ein Grund, weshalb Zuchtwerte nicht über Populationen hinweg vergleichbar sind. Da sich gewisse Rassen weltweit in Teilpopulationen entwickeln, wird von einer Organisation namens Interbull die Umrechnung der geschätzten Zuchtwerte zwischen den verschiedenen Populationen als Dienstleistung angeboten.

Zusammenfassend können wir die folgenden Zwecke der Zuchtwertschätzung identifizieren.

1. Der hauptsächliche Zweck besteht in der Rangierung von Selektionskandidaten nach ihren Zuchtwerten, damit die besten unter ihnen ausgewählt werden können.
2. Geschätzte Zuchtwerte sind auch wichtig für die Bedeutung für die Vorhersage des Zuchtfortschritts, welcher in der Zuchtplanung gebraucht wird.
3. Die Resultate der Zuchtwertschätzung sind auch ein Verkaufsargument für Samendosen von männlichen Tieren. Sie können deshalb auch als Verbraucherinformation betrachtet werden.

Definition 3.1 des Zuchtwerts ist nur schwer in die Praxis umsetzbar, weil man nicht lange warten möchte bis die Anzahl Nachkommen eines Tieres genügend gross ist. In der praktischen Zuchtwertschätzung besteht immer der Zielkonflikt, dass die geschätzten Zuchtwerte möglichst *früh* und auch möglichst *genau* vorliegen sollen. Zur Auflösung dieses Konflikts bestehen verschiedene Methoden. Diese werden im Folgenden vorgestellt werden.

#### 3.4.1.1 Selektionsindex

Beim Selektionsindex sollen die verfügbaren Informationsquellen so kombiniert werden, dass die Korrelation zwischen dem wahren Zuchtwert und dem konstruierten Index maximiert wird. Unter Informationsquellen versteht man alle Leistungen von Tieren, welche aufgrund der Verwandtschaft der Tiere mit dem Probanden zur Schätzung des Zuchtwerts des Probanden beitragen kann. Die Kombination der Informationsquellen wird mit der Konstruktion eines Indexes umgesetzt. Ein Index ( $I$ ) entspricht einer linearen Kombination aus den Informationsquellen (Vektor  $x$ ).

$$I = b_1x_1 + b_2x_2 + \dots + b_nx_n = b^T x \quad (3.5)$$

wobei

- $I$  konstruierter Index
- $b$  Vektor der Gewichte
- $x$  Vektor der Informationsquellen

Statistisch gesehen, entspricht der Selektionsindex, wie er in (3.5) dargestellt ist, einer *multiplen linearen Regression*. Die Regressionskoeffizienten ( $b_i$ ) sind abhängig von

- der Erbllichkeit des Merkmals
- der Art der Informationsquelle
- der additiv-genetischen Verwandtschaft zwischen Probanden und Informationsquellen.

Betreffend der **Terminologie** muss an diesem Punkt nochmals betont werden, dass der **Selektionsindex** (gemäß der Formel (3.5)) zur Schätzung eines Zuchtwertes von Probanden für *ein bestimmtes Merkmal* verwendet wird. Die Informationsquellen bestehen häufig aus phänotypischen Leistungen vom Probanden selber oder von verwandten Tieren. Die Informationen (Vektor  $x$ ) müssen auch nicht notwendigerweise vom gleichen Merkmal sein, wie dasjenige für das wir den Zuchtwert schätzen wollen. Als Beispiel wird in der Schweinezucht oft die Rückenspeckdicke als Informationsquelle für die Schätzung des Zuchtwerts des Merkmals Fleischigkeit verwendet. Diese Verwendung des Begriffs *Selektionsindex* hat nur noch eine historische Bedeutung und ist heute nicht mehr gebräuchlich, da heute andere Verfahren (BLUP-Tiermodell) für die Zuchtwertschätzung eingesetzt werden. Aktuell wird der Begriff des Selektionsindex viel häufiger für die Schätzung des *Gesamtzuchtwerts* verwendet. Die Bedeutung des Begriffs Gesamtzuchtwert wird im Abschnitt 3.4.1.2 erklärt werden.

Zurück zur Schätzung des Zuchtwerts aufgrund des Selektionsindex. Wie wir im Abschnitt 3.3 gesehen hatten, müssen die Informationsquellen in einem ersten Schritt relativiert werden. Hier gehen wir davon aus, dass die Werte  $x_i$  im Vektor  $x$ , bereits Abweichungen von geeigneten Vergleichswerten darstellen. Wir beschäftigen uns im folgenden Abschnitt mit der Berechnung der Faktoren  $b_i$  im Vektor  $b$ .

### 3.4.1.2 Gesamtzuchtwert

Wir beginnen mit dem einfachen und nicht sehr realistischen Fall, dass das Zuchtziel nur aus einem Merkmal besteht. Die verschiedenen verfügbaren Informationsquellen werden zu einem Index  $I$  kombiniert, wobei das Zuchtziel dem wahren Zuchtwert dieses einen Merkmals entspricht und dieser wahre Zuchtziel soll mit  $I$  möglichst genau geschätzt werden. In der praktischen Tierzucht sollen aber meist mehrere Merkmale gleichzeitig verbessert werden. In diesem Fall sprechen wir von einem *komplexen Zuchtziel*. Bei komplexen Zuchtzielen entspricht die Zielgröße der Zuchtwertschätzung einer linearen Funktion ( $T$ ) der wahren Zuchtwerte  $u_i$

$$T = a_1u_1 + a_2u_2 + \dots + a_nu_n = a^T u \quad (3.6)$$

wobei

- $T$  Gesamtzuchtwert
- $a$  Vektor der wirtschaftlichen Gewichte
- $u$  Vektor der wahren Zuchtwerte

Die in Formel (3.6) gezeigte Funktion  $T$  wird als **Gesamtzuchtwert** oder auch als aggregierter Genotyp bezeichnet. Der Vektor  $a$  enthält die *wirtschaftlichen Gewichte*, welche dem Grenznutzen bei einer kleinen Änderung des Merkmals entsprechen. An dieser Stelle nehmen wir an, dass die Werte im Vektor  $a$  bekannt sind. Der Vektor  $u$  enthält die wahren Zuchtwerte.

In einem komplexen Zuchtziel wird  $T$  auf einer monetären Basis, d.h. in einer Währungseinheit ausgedrückt. Hat das Zuchtziel aber nur ein Merkmal, wird die Multiplikation mit den Grenznutzen oft unterlassen. Dann wird der Gesamtzuchtwert  $T$  direkt dem wahren Zuchtwert  $u$  gleichgesetzt.

### 3.4.1.3 Formen von Selektionsindizes

Zusammengefasst haben wir es in der Selektionsindextheorie mit zwei linearen Funktionen zu tun, welche genau getrennt werden müssen.

1. Der eigentliche Selektionsindex ( $I$ ), der den geschätzten Zuchtwert darstellt
2. Der Gesamtzuchtwert ( $T$ ), der eine lineare Kombination der wahren Zuchtwerte des Probanden darstellt

Aufgrund dieser beiden Definitionen können die folgenden Fälle unterschieden werden.

- Unterschiedliche Informationsquellen (Eigenleistung, Geschwister, Nachkommen) aber nur ein Merkmal im Zuchtziel
- Eine Informationsquelle (Eigenleistung), aber mehrere Merkmale im Zuchtziel
- Mehrere Informationsquellen und mehrere Merkmale im Zuchtziel

Die folgenden Umstände können die Zuchtwertschätzung noch zusätzlich erschweren.

- Informationsquellen können auch wiederholte Leistungen des gleichen Merkmals darstellen. Beispiele sind Töchter des gleichen Vaters oder Probewägungen bei der Milchleistung oder die Wurfgrösse bei der gleichen Sau.
- Bei gewissen Merkmalen im Zuchtziel ist die Erhebung sehr umständlich oder sehr teuer. In diesem Fall wird oft auf so genannte Hilfsmerkmale ausgewichen. Das Beispiel der Rückenspeckdicke als Hilfsmerkmal für die Fleischigkeit beim Schwein hatten wir schon früher erwähnt.

### 3.4.1.4 Theorie der Indexkonstruktion

Unter dem Begriff *Indexkonstruktion* versteht man die Berechnung der Gewichtungsfaktoren  $b_i$  aus der Gleichung (3.5). Unabhängig von den Informationsquellen müssen für die Konstruktion die folgenden Parameter bekannt sein.

- Erblichkeiten und phänotypische Standardabweichungen der Merkmale im Zuchtziel und im Index.
- Phänotypische Korrelationen zwischen den Indexmerkmalen
- genetische Korrelationen zwischen Zuchtziel- und Indexmerkmalen
- genetische Korrelationen zwischen Zuchtzielmerkmalen
- ökonomische Gewichte für Zuchtzielmerkmale

Der Index wird mit dem Ziel konstruiert, dass die Korrelation ( $r_{TI}$ ) zwischen dem wahren und geschätzten Zuchtwert maximiert wird. Da der Index ( $I$ ) einer multiplen linearen Regression entspricht, wobei die gesuchten Gewichtungsfaktoren  $b_i$  den Parametern der Regression entsprechen und diese mit Least Squares geschätzt werden, gilt es folglich die mittlere quadrierte Abweichung zwischen den wahren Zuchtwerten und dem Index zu minimieren.

$$\hat{b} = \operatorname{argmin}_b \{E(T - I)^2\} \quad (3.7)$$

Die Gleichung (3.7) bedeutet, dass wir denjenigen Vektor von Indexgewichten  $b$  als Lösung  $\hat{b}$  verwenden, der den Erwartungswert (E) der quadrierten Abweichung zwischen  $T$  und  $I$  minimiert. Für die hier gezeigte Herleitung verwenden wir den einfachen Fall, dass das Zuchtziel nur aus einem Merkmal besteht und der Index nur eine Eigenleistung enthält. Somit gilt

$$T = u \text{ und } I = bx \quad (3.8)$$

Stellen wir daraus die zu minimierende Zielfunktion aus Gleichung (3.7) auf so erhalten wir

$$\begin{aligned}
E(T - I)^2 &= E(u - bx)^2 \\
&= E(u^2 - 2bux + b^2x^2) \\
&= E(u^2) - 2bE(ux) + b^2E(x^2) \\
&= \sigma_u^2 - 2b\sigma_{ux} + b^2\sigma_x^2
\end{aligned} \tag{3.9}$$

Im Schritt von der dritten zur vierten Zeile verwenden wir, dass sowohl  $u$  und  $x$  als Abweichungen definiert sind. Aus diesem Grund dürfen die Erwartungswerte der Quadrate durch die Varianzen und der Erwartungswert des Produkts durch die Kovarianz ersetzt werden.

Zur Minimierung der Zielfunktion aus (3.7) muss die erste Ableitung von (3.9) 0 gesetzt werden<sup>2</sup>. Somit erhalten wir

$$\begin{aligned}
0 &= -2\sigma_{ux} + 2b\sigma_x^2 \\
2b\sigma_x^2 &= 2\sigma_{ux} \\
b &= \frac{\sigma_{ux}}{\sigma_x^2} \\
&= \frac{\sigma_u^2}{\sigma_x^2} = h^2
\end{aligned} \tag{3.10}$$

Schon früher hatten wir gesehen, dass die Regression des Zuchtwerts auf den Phänotyp bei einem Individuum zu einem Regressionskoeffizienten, welcher der Erblichkeit ( $h^2$ ) entspricht, führt.

### 3.4.1.5 Verallgemeinerung

Für die Verallgemeinerung des gezeigten Prinzips braucht es noch die folgende neue Notation.

$\sigma_{pxpy}$	phänotypische Kovarianz zwischen Informationsquellen $x$ und $y$
$\sigma_{pxgy}$	genetische Kovarianz zwischen Informationsquellen $x$ und Zuchtzielmerkmal $y$
$\sigma_{gxgy}$	genetische Kovarianz zwischen Zuchtzielmerkmalen $x$ und $y$

Varianzen von Merkmalen sind als Kovarianzen eines Merkmals mit sich selber in dieser Notation mit berücksichtigt. Analog zur zweiten Zeile in Gleichung (3.10) können wir das Gleichungssystem zur Indexkonstruktion aufschreiben als

$$\begin{aligned}
\sigma_{p1p1}b_1 + \sigma_{p1p2}b_2 + \dots + \sigma_{p1pn}b_n &= \sigma_{p1g1}a_1 + \sigma_{p1g2}a_2 + \dots + \sigma_{p1gm}a_m \\
\sigma_{p2p1}b_1 + \sigma_{p2p2}b_2 + \dots + \sigma_{p2pn}b_n &= \sigma_{p2g1}a_1 + \sigma_{p2g2}a_2 + \dots + \sigma_{p2gm}a_m \\
\dots &= \dots \\
\sigma_{pnp1}b_1 + \sigma_{pnp2}b_2 + \dots + \sigma_{pnpn}b_n &= \sigma_{png1}a_1 + \sigma_{png2}a_2 + \dots + \sigma_{pngm}a_m
\end{aligned} \tag{3.11}$$

Verwandeln wir das Gleichungssystem (3.11) in Matrix-Vektor-Schreibweise, wobei wir  $\sigma_{gipj} = G_{ij}$  und  $\sigma_{pipj} = P_{ij}$  setzen dann folgt

$$Pb = Ga \tag{3.12}$$

Durch Linksmultiplikation beider Seiten in (3.12) mit der Inversen  $P^{-1}$  der Matrix  $P$  ergibt sich die folgende Lösung für den Vektor  $b$

<sup>2</sup>Die Überprüfung der höheren Ableitung, dass es sich beim gefundenen Extremum auch wirklich um ein Minimum handelt, wird hier weggelassen

$$\begin{aligned} P^{-1}Pb &= P^{-1}Ga \\ b &= P^{-1}Ga \end{aligned} \quad (3.13)$$

Da die Matrix  $P$  eine Varianz-Kovarianzmatrix ist, ist sie sicher positiv-definit und somit ist garantiert, dass die Inverse  $P^{-1}$  existiert.

Als Mass der Genauigkeit wird oft die Korrelation zwischen Index und Gesamtzuchtwert verwendet.

$$r_{TI} = \frac{\sigma_{TI}}{\sigma_I \sigma_T} \quad (3.14)$$

Setzen wir dafür die Grössen unseres einfachen Beispiels aus der Herleitung ein dann folgt

$$\begin{aligned} \sigma_T^2 &= E(u^2) = \sigma_u^2 \\ \sigma_I^2 &= E(b^2 x^2) = b^2 E(x^2) = b^2 \sigma_x^2 \\ \sigma_{TI} &= E(u * bx) = b * E(ux) = b \sigma_u^2 \end{aligned} \quad (3.15)$$

Setzen wir (3.15) in die Gleichung (3.14) ein so folgt

$$\begin{aligned} r_{TI} &= \frac{\sigma_{TI}}{\sigma_I \sigma_T} \\ &= \frac{b \sigma_u^2}{b \sigma_x * \sigma_u} \end{aligned} \quad (3.16)$$

Die Korrelation  $r_{TI}$  wird dann maximal<sup>3</sup>, wenn wir  $\sigma_{TI} = \sigma_I^2$  setzen. Somit gilt dann, dass

$$r_{TI} = \frac{\sigma_I^2}{\sigma_I \sigma_T} = \frac{\sigma_I}{\sigma_T} \quad (3.17)$$

Diese Beziehung gilt auch allgemein. Die Korrelation zwischen wahren und geschätztem Zuchtwert entspricht dem Verhältnis der Standardabweichung des geschätzten Zuchtwertes geteilt durch die Standardabweichung des wahren Zuchtwertes. In der Tierzucht wird oft das Quadrat der Genauigkeit  $r_{TI}$  verwendet und als **Bestimmtheitsmass** des Zuchtwertes bezeichnet.

### 3.4.2 Nachkommenleistungen

Die Zuchtwertschätzung aufgrund von Nachkommenleistungen ist immer noch eine der häufigsten Arten, wie Zuchtwerte in der Tierzucht ermittelt werden. Die Nachkommen eines Vatertiers sind oft Halbgeschwister (oder eine Kombination aus Halb- und Vollgeschwister). Die Relativierung der phänotypischen Leistungen wird durch die Korrektur der Leistungen der Nachkommen mit einem geeigneten Vergleichswert ermittelt. Die phänotypische Varianz eines Durchschnitts aus  $n$  vergleichbaren Leistungen ist allgemein

$$\sigma_{\bar{x}}^2 = \frac{1 + (n-1)t}{n} \sigma_x^2 \quad (3.18)$$

wobei  $t$  der Wiederholbarkeit oder Intraklassenkorrelation entspricht. Den Wert für  $t$  kann aus den folgenden Überlegungen hergeleitet werden.

<sup>3</sup>Dies war eine der Anforderungen bei der Konstruktion des Indexes

- Wir nehmen an, dass die  $n$  Tiere untereinander Halbgeschwister sind.
- Halbgeschwister haben im Schnitt  $1/4$  der Allele gemeinsam
- Die Kovarianz zwischen Halbgeschwister beträgt somit  $1/4\sigma_u^2$
- Die Korrelation entspricht der Kovarianz dividiert durch die phänotypische Varianz

Somit gilt für die Halbgeschwister

$$t = 1/4 * \frac{\sigma_u^2}{\sigma_x^2} = \frac{1}{4} * h^2$$

Damit erhalten wir die folgende Gleichung<sup>4</sup> aus der Indexkonstruktion

$$\frac{1 + (n-1)\frac{1}{4} * h^2}{n} * \sigma_x^2 * b = r_{1\alpha} * \sigma_u^2 \quad (3.19)$$

wobei  $r_{1\alpha}$  dem Verwandtschaftsgrad zwischen dem Probanden und dem Nachkommen entspricht. Zwischen Vater und Nachkommen beträgt dieser  $r_{1\alpha} = 1/2$ , da ein Nachkommen die Hälfte der Allele vom Vater erbt. Setzen wir das ein, dann erhalten wir

$$\frac{1 + (n-1)\frac{1}{4} * h^2}{n} * \sigma_x^2 * b = \frac{1}{2} * \sigma_u^2 \quad (3.20)$$

Aufgelöst nach  $b$  ergibt sich das folgende Resultat

$$b = \frac{2nh^2}{4 + (n-1)h^2} \quad (3.21)$$

Setzen wir  $k = (4 - h^2)/h^2$  so erhalten wir für  $b$  den aus der Literatur bekannten Wert von

$$b = \frac{2n}{n+k} \quad (3.22)$$

Die Genauigkeit der so geschätzten Zuchtwerte ergibt sich als

$$r_{TI} = \sqrt{\frac{n}{n+k}} \quad (3.23)$$

Für grosse Werte von  $n$ , d.h. wenn ein Vater sehr viele Nachkommen hat, dann strebt der Wert von  $b$  gegen 2 und erreicht so den theoretischen Wert aus der Definition 3.1. Die Genauigkeit strebt gegen den Wert 1 für sehr grosse Werte von  $n$ . Dies erwarten wir auch intuitiv, dass die Genauigkeit mit zunehmender Anzahl von Nachkommen ansteigt.

### 3.5 Anhang 1: Least Squares Lösungen für die Betriebe

Die Least Squares Lösungen der Betriebe werden aus dem einfachen Regressionsmodell mit dem Betrieb als einzigen fixen Effekt für die Beobachtungen. Das Modell für eine Beobachtung  $y_{ij}$  lautet somit

$$y_{ij} = \beta_j + \epsilon_{ij} \quad (3.24)$$

Die für das Relativieren gebrauchten Least Squares Lösungen der Betriebe erhalten wir mit der Funktion `lm()` in R. Dazu müssen wir zuerst einen dataframe vorbereiten, welche die Zunahmen der Nachkommen und

<sup>4</sup>Diese werden auch als Normalgleichungen bezeichnet

die zugehörigen Betriebe enthält. Wir nennen den dataframe `tbl_beef_farm` und dieser sieht als Tabelle, wie folgt aus.

Betrieb	Zunahme
1	1.26
1	1.32
1	1.40
1	1.44
2	1.52
2	1.50
2	1.42
2	1.46
1	1.34
1	1.32
1	1.24
1	1.28
2	1.44
2	1.40
2	1.54
2	1.56

Bevor wir das Modell in `lm` angeben ist es wichtig, dass wir sicherstellen, dass die Variable `Betrieb` im dataframe `tbl_beef_farm` auch wirklich den Datentyp `factor` hat. Dies können wir mit folgendem Statement überprüfen.

```
is.factor(tbl_beef_farm$Betrieb)
```

```
## [1] TRUE
```

Falls die oben gezeigte Überprüfung nicht das Resultat `TRUE` zurückgibt, dann müssen wir die Variable `Betrieb` in den Datentypen `factor` umwandeln mit dem Befehl

```
tbl_beef_farm$Betrieb <- as.factor(tbl_beef_farm$Betrieb)
```

Das Modell wird dann spezifiziert mit

```
lm_beef_farm <- lm(Zunahme ~ 0 + Betrieb, data = tbl_beef_farm)
```

Die Resultate der Anpassung des Regressionsmodells erhalten wir mit der `summary()`-Funktion

```
summary(lm_beef_farm)
```

```
##
## Call:
## lm(formula = Zunahme ~ 0 + Betrieb, data = tbl_beef_farm)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.08500 -0.04875 -0.00500  0.04500  0.11500
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## Betrieb1  1.32500     0.02248   58.94  <2e-16 ***
## Betrieb2  1.48000     0.02248   65.84  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06358 on 14 degrees of freedom
## Multiple R-squared:  0.9982, Adjusted R-squared:  0.998
```

```
## F-statistic: 3904 on 2 and 14 DF, p-value: < 2.2e-16
```

Die Least Squares Lösungen für die Betriebe erhalten wir mit dem Statement

```
coefficients(lm_beef_farm)
```

```
## Betrieb1 Betrieb2
```

```
## 1.325 1.480
```