

# Züchtungslehre - Tierzucht und Selektionsindex

*Peter von Rohr*

*2016-10-07*

## Tierzucht

In der Tierzucht gibt es zwei **fundamentale Fragen**, welche alle Züchterinnen und Züchter beschäftigt.

1. Welches Tier ist das beste?
2. Wie sollen Tiere ausgewählt werden, dass ihre Nachkommen besser sind als die aktuelle Population

## Das beste Tier

Da wir im allgemeinen nicht nur an einer Eigenschaft bei den Nutztieren interessiert sind, stellt sich die Frage nach der optimalen Kombination an Eigenschaften. Ist zum Beispiel die Milchkuh die beste, welche eine möglichst hohe Milchleistung hat oder spielen Eutergesundheit, ein gutes Fundament oder Krankheitsresistenz auch eine Rolle? Müssen Schweine möglichst schnell wachsen oder sind ein gutes Geburtsverhalten und überdurchschnittliche Fleischqualität auch wichtig?

Die Bezeichnung “das beste Tier” ist ein relativer Begriff. Absolut gesehen gibt es kein bestes Tier. Die Klassierung von Tieren ist immer relativ zu einer Umwelt, in welcher sich die Tiere befinden. Eine Hochleistungsmilchkuh hätte sehr viel Mühe in einem Wüstengebiet in Afrika zu überleben. Bei den kommenden Ausführungen gehen wir davon aus, dass die Umwelt für die Tiere optimal ist.

Die Beschreibung von Tieren basiert entweder auf deren äusserem Erscheinungsbild oder aufgrund von Leistungen oder einer Kombination von beidem. In jedem Fall sprechen wir von **Merkmalen**. Ein Merkmal ist definiert als beobachtbare oder messbare Eigenschaft eines Tieres.

Als Beispiele für beobachtete Merkmale sind

- Fellfarbe
- Grösse
- Bemuskelung
- Fundament
- Euter

Beispiele für messbare Merkmale sind

- Körpergewicht
- Milchproduktion
- Wachstum
- Futterraufnahme

Wichtig ist, dass es bei den Merkmalen nicht um die Beschreibung eines einzelnen Tieres geht. Ein Tier kann rote Fellfarbe und ein Körpergewicht von  $343\text{kg}$  aufweisen. Aber **rot** und  $343\text{kg}$  sind nicht die Merkmale, sondern die Merkmalsausprägungen oder der **Phänotyp**. Die Merkmale sind nur **Körpergewicht** und **Fellfarbe**.

In der Tierzucht beschäftigen wir uns vor allem mit der genetischen Veränderung von Populationen. Das genetische Potential wird von einem Tier an die nächste Generation weitergegeben. Das genetische Potential

eines Tieres wird durch den **Genotypen** repräsentiert. Den Zusammenhang zwischen phänotypischem Merkmal und Genotyp eines Individuums wird durch das folgende uns schon bekannte Modell hergestellt.

$$y = \mu + g + \epsilon$$

wobei  $y$  der phänotypische Wert eines Merkmals ist,  $\mu$  dem Populationsmittel entspricht,  $g$  für den genotypischen Wert steht und  $\epsilon$  der umweltbedingte Rest darstellt.

Schon im Kapitel "Quantitative Genetik" hatten wir gesehen, dass aufgrund der Vererbung, nicht der Genotyp eines Tieres an die nächste Generation weitergegeben wird sondern nur eine zufällige Auswahl der Allele. Somit ist nicht der genotypische Wert sondern der genetisch additive Anteil des genotypischen Wertes von zentraler Bedeutung. Den genetisch additiven Anteil hatten wir als **Zuchtwert** bezeichnet und wir sehen hier, dass der Zuchtwert eines Tieres in der Beurteilung der Tiere hinsichtlich ihrer Eignung als Elterntiere eine zentrale Rolle spielt. Formell hatten wir im Kapitel "Quantitative Genetik" den genotypischen Wert  $g$  in Zuchtwert  $a$ , Dominanzabweichung  $d$  und Epistasie  $i$  aufgeteilt. Somit setzt sich der phänotypische Wert eines Merkmals wie folgt zusammen.

$$y = \mu + a + d + i + \epsilon$$

Bei der Dominanzabweichung  $d$  und der Epistasie  $i$  handelt es sich um Interaktionen an einem Locus oder zwischen Loci. Diese sind somit abhängig vom Paarungspartner und nur schwer zu quantifizieren. Deshalb werden diese beiden Effekte auch häufig zusammen mit den Umweltfaktoren  $e$  in einen neuen Restterm  $e$  zusammengefasst.

$$y = \mu + a + e \tag{1}$$

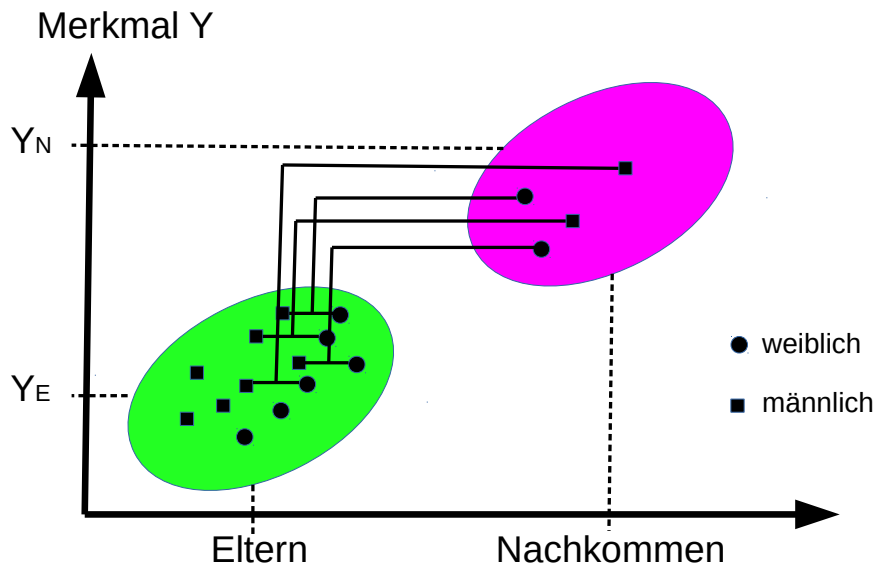
### Auswahl der Elterntiere

Bei der Auswahl der Elterntiere stehen uns zwei Werkzeuge zur Verfügung.

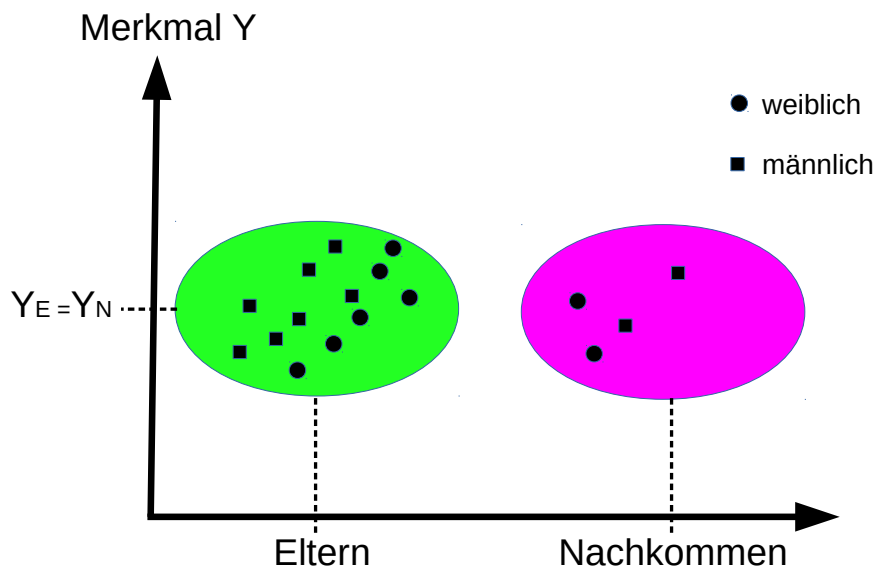
1. Selektion
2. Gezielte Anpaarung

### Selektion

Selektion wird verwendet um langfristige genetische Änderungen in einer Population zu bewirken. Der Unterschied zwischen gerichteter Selektion und Vermehrung sind in den nachfolgenden beiden Diagrammen gezeigt.



Im Gegensatz zur gerichteten Selektion steht die reine Vermehrung, welche im folgenden Schema gezeigt ist. Bei der Vermehrung wird keine Verbesserung der Nachkommen im Vergleich zu den Elterntieren angestrebt.



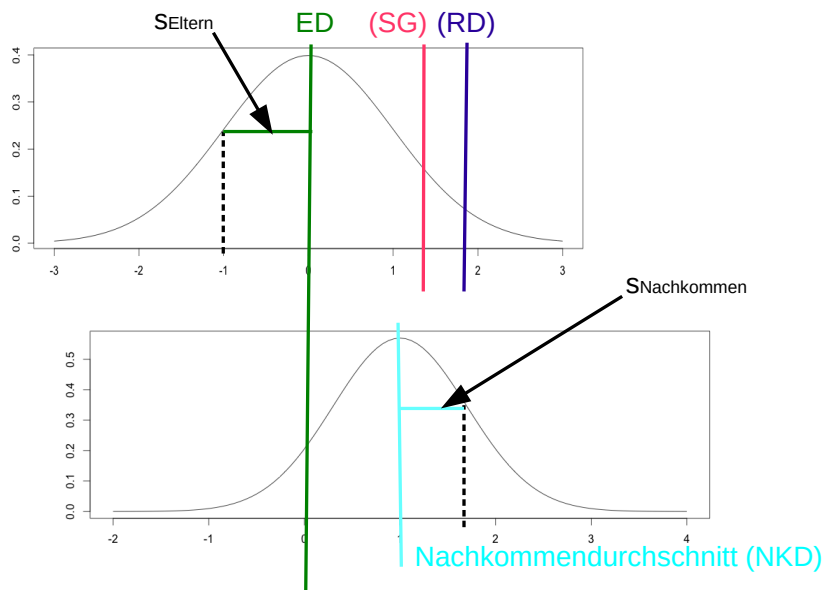
Als Selektion versteht man den Prozess, welcher bestimmt welche Tiere aus einer Population Eltern werden, wie viele Nachkommen sie produzieren sollen und wie lange sie in der Zuchtpopulation verbleiben. Wir sind vertraut mit der **natürlichen Selektion**, welche auch oft mit dem Schlagwort “survival of the fittest” zusammengefasst wird. Damit ist gemeint, dass Tiere, welche optimal an eine Umwelt angepasst sind haben eine höhere Wahrscheinlichkeit zu überleben und sich somit fortzupflanzen.

In der Tierzucht ist aber vor allem die **künstliche Selektion** von Interesse. Die Idee hinter der künstlichen

Selektion ist, die Individuen mit dem besten Set von Genen auszuwählen und diese sich vermehren zu lassen. Die Tiere mit dem besten Set von Genen haben den besten Wert als Zuchttiere. Dieser Wert wird anhand unseres vorher definierten Begriffs des **Zuchtwertes** gemessen.

Die einfachste Form der Selektion ist die phänotypische Selektion. Dabei werden die Tiere alleine aufgrund der phänotypischen Leistung selektiert. Diese Art der Selektion wird kaum mehr verwendet, ist aber sehr einfach und illustrativ für Erklärungen. Aktuell sollten Tiere aufgrund ihrer **Zuchtwerte** selektiert werden. Unter der Annahme, dass Zuchtwerte normal verteilt sind, sind die Vorgänge von der phänotypischen Selektion auf die Selektion nach Zuchtwerten übertragbar.

Das nachfolgende Bild zeigt die Verteilung eines phänotypischen Merkmals in einer Population. Dabei stehen ED für Elterndurchschnitt, SG für Selektionsgrenze und RD für Remontendurchschnitt.



## Gezielte Paarung

Die gezielte Paarung ist der Prozess, in welchem bestimmt wird, welche Tiere miteinander gepaart werden sollen. Dies ist verschieden von der Selektion. In der Selektion werden nur die zukünftigen Eltern aus der Population ausgewählt. Bei der gezielten Paarung geschieht die Zuordnung der Paarungspartner. Die Zuordnung der Paarungspartner kann auf sehr verschiedene Arten passieren. Die Art der Zuordnung bestimmt dann das Paarungssystem. Dabei können verschiedene Ziele verfolgt werden, wie zum Beispiel

- a) Maximierung des Zuchtwertes
- b) Ausnützung gegensätzlicher Eigenschaften
- c) Erzielung von Heterosis-Effekten

Paarungen können innerhalb der gleichen Rasse geplant werden, dann spricht man von **Reinzucht**. Werden Partner aus verschiedenen Rassen gepaart, bezeichnen wir das als **Kreuzungszucht**.

Der Mittelwert der Zuchtwerte der Elterntiere liefert uns eine erste Information über den zu erwartenden Zuchtwert des Nachkommens. Dieser wird als **Abstammungszuchtwert** bezeichnet. Dieser erste geschätzte Zuchtwert liegt schon vor bevor der Nachkomme geboren wurde. Abstammungszuchtwerte sind aber nicht die einzige Art von Zuchtwerten, welche in der Tierzucht verwendet werden. In den nachfolgenden Abschnitten

werden wir uns noch etwas genauer damit beschäftigen, welche verschiedenen Arten von Zuchtwerten es gibt und wie diese geschätzt werden.

## Zuchtwertschätzung unter Verwendung verschiedener Informationsquellen

In jedem Zuchtprogramm ist die Schätzung von Zuchtwerten ein zentraler Bestandteil. Die Genauigkeit der geschätzten Zuchtwerte hängt sehr stark von den verfügbaren Daten ab. Bei den Daten handelt es sich meist um phänotypische Beobachtungswerte und um Verwandtschaftsbeziehungen. Die verfügbaren SNP-Informationen für die genomische Selektion wird hier noch nicht betrachtet.

Zu Beginn der züchterischen Arbeit sind meist nur phänotypische Informationen von einzelnen Tieren verfügbar. In späteren Stadien kommen Messungen von verwandten Tieren und von Nachkommen hinzu. Zuchtwerte müssen also zuerst auf der Basis von phänotypischen Informationen des Tieres alleine geschätzt werden. Nach und nach können für die Zuchtwertschätzung auch Informationsquellen von verwandten Tieren berücksichtigt werden.

Meist werden in Zuchtprogrammen nicht nur einzelne Merkmale bearbeitet. Häufig sind ganze Gruppen von Merkmalen von Interesse. Verschiedene Merkmale können gleichzeitig über die Definition eines sogenannten **Selektionsindex** berücksichtigt werden.

### Das Modell

Eine phänotypische Beobachtung  $y_{ij}$  eines Tieres  $i$  in einem Merkmal  $j$  ist beeinflusst durch allgemeine identifizierbare Umweltfaktoren  $\mu_i$ , durch die additiv genetischen Effekte  $a_i$  und durch einen Restterm  $e_{ij}$ .

$$y_{ij} = \mu_i + a_i + e_{ij} \quad (2)$$

Die Aufteilung des totalen genotypischen Wertes  $g_i$  in additiven Effekt  $a_i$ , Dominanzabweichung  $d_i$  und Epistasie  $i_i$  und die Verschiebung der Dominanzabweichung und der Epistasie in den Restterm wurde schon bei der Herleitung des Modells (1) besprochen und deshalb hier nicht weiter ausgeführt.

Das Modell in (2) ist ein lineares Modell, welches in der Tierzucht sehr weit verbreitet ist. Meist wird angenommen, dass der Vektor  $y$  aller Beobachtungen in einer Population einer multivariaten Normalverteilung folgt. Diese Annahme wird durch den zentralen Grenzwertsatz begründet, welcher besagt, dass eine Zufallsvariable, welche aus einer Summe mit sehr vielen Summanden mit sehr kleinem Effekt zusammengesetzt ist, sich einer Normalverteilung nähert. Das Argument des zentralen Grenzwertsatzes wird zuerst einmal auf den additiv genetischen Effekt  $a$  angewendet. Auch für den Resteffekt  $e$  wird angenommen, dass sich diese zufälligen Resteffekte aus sehr vielen kleinen Effekten zusammensetzen. Somit folgen  $a$  und  $e$  einer Normalverteilung. Die systematischen mittleren Umwelteffekte  $\mu$  werden als fix angenommen und zeigen somit nicht den Charakter einer Zufallsvariablen mit einer Verteilung. Da  $a$  und  $e$  normal verteilt sind und  $\mu$  als fixer Effekt angenommen wird, folgt auch  $y$  einer Normalverteilung. Dieses hier beschriebene Modell trägt den Namen **infinitesimal model**.

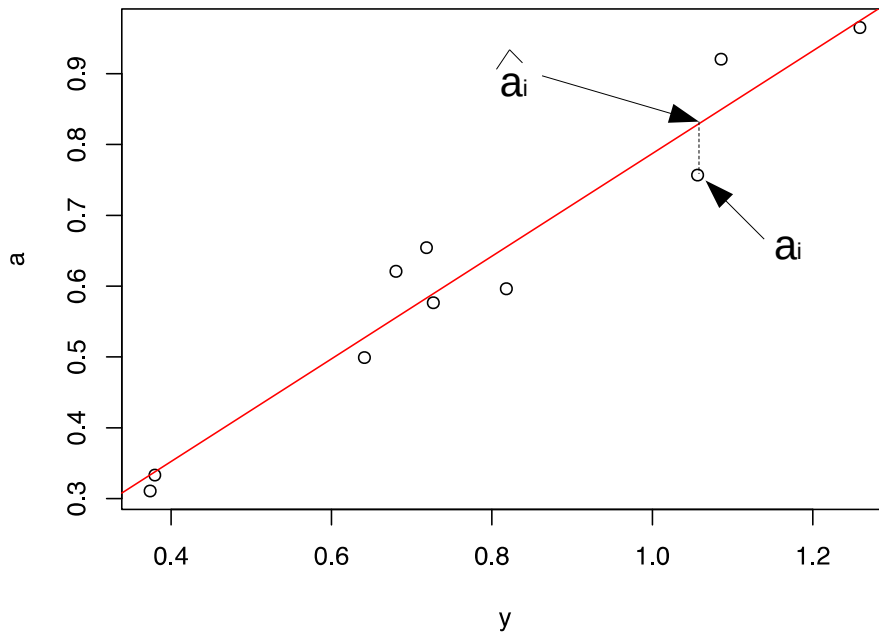
Zu den Annahmen der Verteilungen der Zufallsvariablen im Modell (2) gehören auch die **Erwartungswerte** und die **Varianzen**. Da die Effekte  $a$  und  $e$  als Abweichungen von einem Mittel definiert sind, ist deren Erwartungswert  $E(a) = E(e) = 0$ . Somit gilt für den Erwartungswert  $E(y)$  von  $y$ :  $E(y) = \mu$ . Die Varianzen von  $a$  und  $e$  werden als bekannt angenommen. Zwischen  $a$  und  $e$  sollen keine Kovarianzen auftreten, d.h.  $cov(a, e) = 0$ . Auch zwischen den Resteffekten verschiedener Tiere  $i$  und  $k$  sollen keine Kovarianzen bestehen, d.h.  $cov(e_i, e_k) = 0$ . Aus diesen Annahmen lässt sich für die Varianz von  $y$  ableiten, dass

$$var(y) = var(a) + var(e)$$

# Schätzung des Zuchtwertes aufgrund von Eigenleistungen

## Eine einzelne Eigenleistung

Falls nur eine Leistung pro Tier zur Verfügung steht, dann lassen sich die Zuchtwerte aus der Regression der phänotypischen Leistungen auf die additiv genetischen Werte schätzen. Da die additiv genetischen Werte unbekannt sind, ist auch die Regression unbekannt. Aber es ist trotzdem hilfreich sich vorzustellen, wie diese Regression aussehen würde.



Die rote Gerade im obigen Diagramm zeigt die Regressionsgerade der additiv genetischen Werte auf die phänotypischen Beobachtungen. Da wir den additiv genetischen Wert  $a_i$  nicht kennen schätzen wir ihn und bezeichnen den Schätzwert als  $\hat{a}_i$ . Den Schätzwert berechnen wir als

$$\hat{a}_i = b(y_i - \mu) \quad (3)$$

wobei  $b$  der Regressionskoeffizient der additiv genetischen Werte auf die phänotypischen Beobachtungen darstellt. Den Regressionskoeffizienten  $b$  berechnen wir aus den Kovarianzen der Zufallsvariablen als

$$b = \frac{\text{cov}(a, y)}{\text{var}(y)} \quad (4)$$

wobei

$$\text{cov}(a, y) = \text{cov}(a, \mu + a + e) = \text{cov}(a, \mu) + \text{cov}(a, a) + \text{cov}(a, e)$$

Aus den Annahmen zum Modell (2) wissen wir, dass  $\text{cov}(a, \mu) = \text{cov}(a, e) = 0$ . Die Kovarianz der Zufallsvariablen mit sich selber entspricht der Varianz der Zufallsvariablen und somit ist  $\text{cov}(a, a) = \text{var}(a)$ . Somit haben wir als Resultat für den Regressionskoeffizienten

$$b = \frac{\text{cov}(a, y)}{\text{var}(y)} = \frac{\text{var}(a)}{\text{var}(y)} = h^2 \quad (5)$$

Der Term  $h^2$  steht für die Heritabilität. Somit entspricht der geschätzte Zuchtwert aufgrund einer Eigenleistung das Produkt aus Heritabilität mal die um den Mittelwert  $\mu$  korrigierten Phänotypen  $y$

$$\hat{a}_i = b(y_i - \mu) = h^2(y_i - \mu) \quad (6)$$

Die Korrelation zwischen dem wahren Zuchtwert und dem Selektionskriterium, hier die phänotypische Beobachtung, entspricht der Genauigkeit des geschätzten Zuchtwertes. Diese Korrelation bietet eine Möglichkeit die verschiedenen Selektionskriterien auf ihre Genauigkeit hin zu vergleichen. Je höher die Korrelation, desto besser die Genauigkeit und desto geeigneter das Selektionskriterium. Die Schätzgenauigkeit wird sehr oft auch als Bestimmtheitsmass ( $r^2$ ) ausgewiesen. Für unser Beispiel einer einzelnen Eigenleistung ist die Genauigkeit

$$\begin{aligned} r_{a,y} &= \frac{\text{cov}(a, y)}{\sigma_a \sigma_y} \\ &= \frac{\sigma_a^2}{\sigma_a \sigma_y} = h \end{aligned} \quad (7)$$

und das Bestimmtheitsmass  $r^2 = h^2$ .

Der **erwartete Selektionserfolg** ( $R$ ) auf der Basis einer einzelnen Eigenleistung beträgt

$$R = i r_{a,y}^2 \sigma_y = i h^2 \sigma_y \quad (8)$$

wobei  $i$  die Selektionsintensität ist. Diese entspricht der Überlegenheit der selektierten Individuen ausgedrückt in Einheiten einer phänotypischen Standardabweichung.

Die Varianz der geschätzten Zuchtwerte ist

$$\text{var}(\hat{a}) = \text{var}(by) = \text{var}(h^2 y) = h^4 \sigma_y^2 = r_{a,y}^2 h^2 \sigma_y^2 = r_{a,y}^2 \sigma_a^2 \quad (9)$$

## Wiederholte Messungen

Falls bei einem Tiere mehrere wiederholte Messungen des gleichen Merkmals vorliegen, kann der Zuchtwert aus dem Mittelwert der wiederholten Messungen geschätzt werden. Mit den wiederholten Messungen wird angenommen, dass sich die Messungen eines Tieres aufgrund der permanenten Umweltfaktoren ähnlicher sind. Was zwischen den wiederholten Messungen bei einem Tier variiert sind die sogenannten temporären Umwelteffekte. Es gibt also eine zusätzliche Kovarianz bei Messungen eines Tieres. Die Varianz ( $\text{var}(y)$ ) der phänotypischen Beobachtung kann somit aufgeteilt werden in

$$\text{var}(y) = \text{var}(a) + \text{var}(pe) + \text{var}(te)$$

wobei  $\text{var}(pe)$  die Varianz aufgrund der permanenten Umwelt darstellt und  $\text{var}(te)$  der zufälligen Umweltvarianz entspricht.

Neu wird der Begriff der Wiederholbarkeit  $t$  oder intra-klassen Korrelation definiert als

$$t = \frac{\text{var}(a) + \text{var}(pe)}{\text{var}(y)} \quad (10)$$

Bei diesem Modell nehmen wir an, dass die wiederholten Messungen bei einem Tier  $i$  vom gleichen Merkmal her stammen. Somit ist die genetische Korrelation zwischen den wiederholten Messungen gleich eins, da es sich um das gleiche Merkmal handelt. Weiter wird angenommen, dass die Varianzen der Messungen alle gleich sind und dass die Korrelationen der Umwelteffekte zwischen zwei Messungen gleich sind.

Der Zuchtwert  $\hat{a}_i$  soll nun aufgrund des Mittelwertes  $\bar{y}$  der wiederholten Messungen geschätzt werden.

$$\hat{a}_i = b(\bar{y}_i - \mu) \quad (11)$$

Hier wird der Regressionskoeffizient  $b$  berechnet als

$$b = \frac{\text{cov}(a, \bar{y})}{\text{var}(\bar{y})} \quad (12)$$

dabei ist

$$\text{cov}(a, \bar{y}) = \text{cov}(a, a + pe + \frac{1}{n} \sum te) = \sigma_a^2$$

und

$$\text{var}(\bar{y}) = \text{var}(a) + \text{var}(pe) + \frac{1}{n} \text{var}(te)$$

Kombinieren wir das mit der Definition in (10) dann führt das zu

$$\text{var}(\bar{y}) = [t + (1 - t)/n] \sigma_y^2 \quad (13)$$

Setzen wir diese Beziehungen für  $\text{cov}(a, \bar{y})$  und für  $\text{var}(\bar{y})$  in Gleichung (12) ein dann erhalten wir als Resultat für  $b$

$$b = \frac{\sigma_a^2}{[t + (1 - t)/n] \sigma_y^2} = \frac{nh^2}{1 + (n - 1)t}$$

Den Zuchtwert  $\hat{a}_i$  erhalten wir durch Einsetzen von  $b$  in Gleichung (11)

$$\hat{a}_i = \frac{nh^2(\bar{y}_i - \mu)}{1 + (n - 1)t}$$

Die Genauigkeit des geschätzten Zuchtwertes berechnen wir wieder als Korrelation des wahren Zuchtwertes und des Selektionskriteriums.

$$r_{a, \bar{y}} = \frac{\text{cov}(a, \bar{y})}{\sigma_a \sigma_{\bar{y}}} = \frac{\sigma_a^2}{\sigma_a \sigma_y \sqrt{[t + (1 - t)/n]}} = \frac{h\sqrt{n}}{\sqrt{1 + (n - 1)t}} = \sqrt{\frac{nh^2}{1 + (n - 1)t}} = \sqrt{b} \quad (14)$$

Wiederholte Messungen bringen einen Zuwachs an Genauigkeit. Bilden wir das Verhältnis der beiden Genauigkeiten aus (14) und (7) so erhalten wir

$$\frac{r_{a, \bar{y}}}{r_{a, y}} = \frac{\sqrt{\frac{nh^2}{1 + (n - 1)t}}}{h} = \sqrt{\frac{n}{1 + (n - 1)t}}$$

Dieser Ausdruck zeigt, dass zusätzliche Messungen vor allem bei tiefer Wiederholbarkeit einen grossen Zuwachs an Genauigkeit bringen.



Der erwartete Selektionserfolg  $R$  für wiederholte Messungen beträgt

$$R = i * r_{a,\tilde{y}}^2 * \sigma_y \sqrt{[t + (1-t)/n]}$$

### Zuchtwertschätzung aufgrund von Nachkommen

Für Merkmale, welche nur bei weiblichen Tieren beobachtbar sind, werden Zuchtwerte für männliche Tiere aufgrund der durchschnittlichen Leistungen ihrer Nachkommen geschätzt. Diese Situation ist typisch für Milchrinder, bei welchen die Milchleistung nur bei den weiblichen Tieren erhoben werden kann. Zuchtwerte der Bullen werden dann aufgrund der Töchterleistung geschätzt. Wir nehmen also an, dass  $\tilde{y}_s$  der Mittelwert von  $n$  Messungen der Nachkommen von Stier  $s$  darstellen. Die  $n$  Töchter von Stier  $s$  sind also Halbgeschwister und wir nehmen an, dass sie nur über den Vater verwandt sind. Der geschätzte Zuchtwert von Stier  $s$  ist somit

$$\hat{a}_s = b(\tilde{y}_s - \mu) \quad (15)$$

wobei

$$b = \frac{\text{cov}(a, \tilde{y})}{\text{var}(\tilde{y})} \quad (16)$$

Wir berechnen nun Zähler und Nenner von (16) separat und fügen diese dann wieder zusammen. Als Vorbereitung schauen wir uns noch die Zerlegung der phänotypischen Beobachtung an.

$$\tilde{y}_s = \mu + \frac{1}{n} \sum_i a_i + \frac{1}{n} \sum_i e_i$$

Die genetisch additiven Werte  $a_i$  lassen sich zerlegen in

$$a_i = \frac{1}{2}(a_s + a_d) + m_i$$

wobei  $m_i$  als **Mendelian Sampling** bezeichnet wird und die Abweichung eines Zuchtwertes vom Eltern-durchschnitt darstellen. Diese  $m_i$ -Effekte sind unkorreliert zu den Zuchtwerten und heben sich aufgrund ihrer Definition als Abweichungen über alle Halbgeschwister auf, d.h.  $\sum_i m_i = 0$ .

$$\tilde{y}_s = \mu + \frac{1}{2}a_s + \frac{1}{n} \sum_i a_{d(i)}/2 + \frac{1}{n} \sum_i e_i \quad (17)$$

Dies setzen wir in die Formel für die Kovarianz ein. Dabei vereinfacht sich der Ausdruck für die Kovarianz dadurch, dass  $a$  jeweils unkorreliert zu  $\mu$ ,  $m_i$  und  $e$  ist. Da nur die Verwandtschaft über den Vater berücksichtigt wird, ist auch  $a$  und  $a_d$  unkorreliert.

$$\text{cov}(a, \tilde{y}) = \text{cov}(a, \mu + \frac{1}{2}a_s + \frac{1}{n} \sum_i a_{d(i)}/2 + \frac{1}{n} \sum_i e_i) = \text{cov}(a, \frac{1}{2}a_s) = \frac{1}{2} \text{cov}(a, a_s) = \frac{1}{2} \text{var}(a)$$

Für die Varianz  $\text{var}(\tilde{y})$  gehen wir analog vor wie bei den wiederholten Messungen und verwenden die Formel (13).

$$\text{var}(\tilde{y}) = [t + (1-t)/n] \sigma_y^2 \quad (18)$$

Die Intra-Klassen-Korrelation  $t$  zwischen den Messungen der Halbschwestern ist hier

$$t = \frac{0.25\sigma_a^2}{\sigma_y^2} = \frac{1}{4}h^2$$

Aus den bisherigen Zwischenresultaten können wir nun den Regressionskoeffizienten berechnen.

$$\begin{aligned} b &= \frac{\frac{1}{2}\sigma_a^2}{[t + (1-t)/n]\sigma_y^2} \\ &= \frac{\frac{1}{2}h^2\sigma_y^2}{[\frac{1}{4}h^2 + (1 - \frac{1}{4}h^2)/n]\sigma_y^2} \\ &= \frac{\frac{1}{2}nh^2}{\frac{1}{4}nh^2 + (1 - \frac{1}{4}h^2)} \\ &= \frac{2nh^2}{nh^2 + 4 - h^2} \\ &= \frac{2n}{n + \frac{(4-h^2)}{h^2}} \\ &= \frac{2n}{n + k} \end{aligned} \tag{19}$$

wobei  $k = \frac{(4-h^2)}{h^2}$ .

Somit ist der geschätzte Zuchtwert  $\hat{a}_s$  aufgrund eines Durchschnitts von Nachkommenleistungen

$$\hat{a}_s = \frac{2n}{n + k} (\bar{y}_s - \mu) \tag{20}$$

Die Genauigkeit des geschätzten Zuchtwertes  $\hat{a}_s$  berechnet sich als

$$\begin{aligned} r_{a,\bar{y}} &= \frac{cov(a, \bar{y})}{\sqrt{var(a) var(\bar{y})}} \\ &= \frac{\frac{1}{2}var(a)}{\sqrt{var(a) [\frac{1}{4}h^2 + (1 - \frac{1}{4}h^2)/n]\sigma_y^2}} \\ &= \frac{\frac{1}{2}h^2\sigma_y^2}{\sqrt{h^2\sigma_y^2 [\frac{1}{4}h^2 + (1 - \frac{1}{4}h^2)/n]\sigma_y^2}} \\ &= \sqrt{\frac{nh^2}{nh^2 + (4 - h^2)}} \\ &= \sqrt{\frac{n}{n + k}} \end{aligned} \tag{21}$$

### Zuchtwertschätzung aufgrund von Pedigreeinformation

Falls für ein Tier keine phänotypischen Beobachtungen existieren, dann kann sein Zuchtwert aufgrund der geschätzten Elternzuchtwerte geschätzt werden. Dies entspricht dann einem Abstammungszuchtwert. Der geschätzte Zuchtwert  $\hat{a}_o$  für Tier  $o$  aufgrund der geschätzten Zuchtwerte  $\hat{a}_s$  und  $\hat{a}_d$  der Eltern  $s$  und  $d$  ist somit

$$\hat{a}_o = \frac{1}{2}(\hat{a}_s + \hat{a}_d)$$

Setzen wir  $f = \frac{1}{2}(\hat{a}_s + \hat{a}_d)$ , dann ist die Genauigkeit

$$r_{a,f} = \frac{\text{cov}(a_o, f)}{\sqrt{\sigma_a^2 \text{var}(1/2\hat{a}_s + 1/2\hat{a}_d)}}$$

Unter der Annahme, dass die Eltern  $s$  und  $d$  nicht verwandt sind, folgt

$$\text{cov}(a_o, f) = \frac{1}{4}\text{cov}(a_s, \hat{a}_s) + \frac{1}{4}\text{cov}(a_d, \hat{a}_d) = \frac{1}{4}\text{var}(\hat{a}_s) + \frac{1}{4}\text{var}(\hat{a}_d)$$

Verwenden wir für die Varianzen der geschätzten Zuchtwerte die Beziehung in (9) dann folgt

$$\text{cov}(a_o, f) = \frac{1}{4}(r_s^2 + r_d^2)\sigma_a^2$$

Aufgrund der hier gezeigten Berechnungen folgt auch, dass

$$\text{var}(1/2\hat{a}_s + 1/2\hat{a}_d) = \frac{1}{4}(r_s^2 + r_d^2)\sigma_a^2$$

. Somit folgt für die Genauigkeit

$$r_{a,f} = \frac{\frac{1}{4}(r_s^2 + r_d^2)\sigma_a^2}{\sqrt{\sigma_a^2 \frac{1}{4}(r_s^2 + r_d^2)\sigma_a^2}} = \frac{1}{2}\sqrt{(r_s^2 + r_d^2)}$$

## Selektionsindex

Traditionellerweise bezeichnete man die Methode um Zuchtwerte eines Merkmals aufgrund verschiedener phänotypischer Informationsquellen zu schätzen als **Selektionsindex**. Bei dieser Methode wurden die verschiedenen Beobachtungen zu einem sogenannten Index  $I$  zusammengefasst. Diese Zusammenfassung entsprach einem gewichteten Mittel der Abweichungen der phänotypischen Beobachtung von deren Mittelwert. Für ein bestimmtes Tier  $i$  war der Index definiert als

$$I_i = \hat{a}_i = b_1(y_1 - \mu_1) + b_2(y_2 - \mu_2) + b_3(y_3 - \mu_3) + \dots$$

Die Gewichtungsfaktoren  $b$  werden so bestimmt, dass die mittleren quadrierten Fehler der geschätzten Zuchtwerte minimal sind. Diese Art der Schätzung von Zuchtwerten wurde bis etwa in die 1990er-Jahre bei den Schweinen auf der Stationprüfung verwendet. Seit dann findet diese Methode aber keine Anwendung mehr.

## Schätzung des Gesamtzuchtwertes

Die Indextheorie wird aber in einem anderen Gebiet immer noch verwendet und zwar zur Schätzung des Gesamtzuchtwertes. In jedem Zuchtprogramm muss ja ein Zuchtziel vorliegen. Dies ist meist eine Vorstellung der Eigenschaften eines optimalen Tieres. Damit wir mit züchterischen Methoden unser Ziel erreichen können müssen wir eine mathematische Formulierung des Zuchtziels definieren. Diese mathematische Version des Zuchtziels bezeichnen wir als **Gesamtzuchtwert**  $H$ . Da wir in unserem Gesamtzuchtwert mehrere Merkmale berücksichtigen ist auch  $H$  als gewichtetes Mittel der wahren Zuchtwerte der Merkmale im Gesamtzuchtwert definiert. Die Gewichtungsfaktoren entsprechen den wirtschaftlichen Gewichten der Merkmale im Zuchtziel.

Das wirtschaftliche Gewicht  $w$  eines Merkmals entspricht dem zusätzlichen Gewinn, welcher erzielt wird, wenn sich das Populationsmittel eines Merkmals um eine kleine Einheit ändert. Als Beispiel wäre das wirtschaftliche Gewicht für Zuwachs bei Fleischrindern, der zusätzliche Gewinn, den wir erzielen, falls unsere Population eine um 1 Gramm erhöhte Tageszunahme hat.

Somit ist der Gesamtzuchtwert  $H$  definiert als

$$H = w_1 * a_1 + w_2 * a_2 + w_3 * a_3 + \dots$$

Den Gesamtzuchtwert  $H$  können wir nur schätzen, da die wahren Zuchtwerte unbekannt sind. Für die Schätzung von  $H$  stellen wir einen Index auf, welcher jetzt nicht nur einen Zuchtwert sondern den Gesamtzuchtwert  $H$  schätzen soll. Wir definieren also einen neuen Index

$$I = \hat{H} = b_1 * x_1 + b_2 * x_2 + \dots$$

Hier stehen die  $x_i$  für die verfügbaren Informationen. Meist sind dies mit BLUP geschätzte Zuchtwerte (siehe später). Die Merkmale im Gesamtzuchtwert und im Index müssen nicht die gleichen sein. Die Gewichtungsfaktoren  $b_i$  werden so abgeleitet, dass die Varianz  $var(H - I)$  des Fehlers minimal ist. Daraus folgt in Matrix-Vektor-Schreibweise die sogenannte **Indexgleichung**

$$Pb = Gw$$

wobei  $P$  die Co-Varianzmatrix der Merkmale im Index ist und  $G$  die genetische Co-Varianzmatrix zwischen den Merkmalen im Index und den Merkmalen im Gesamtzuchtwert darstellt.

Aufgrund der Indexgleichung können die Gewichtungsfaktoren  $b$  berechnet werden als

$$b = P^{-1}Gw$$