

## 9.6 Nachtrag zur genetisch additiven Verwandtschaftsmatrix

(Hofer, 1990), welcher (Dempfle, 1987) zitiert weist auf zwei unterschiedliche Interpretationen der genetisch additiven **Verwandtschaftsmatrix**  $A$  hin. Die eine basiert auf der Wahrscheinlichkeit, dass zwei Allele herkunftsgleich sind, d.h. zwei Kopien eines und des selben Ahnenallels (IBD) entsprechen. Diese entspricht der Wahrscheinlichkeit, dass an einem beliebigen Locus zwei Allele von zwei Tieren  $s$  und  $d$  herkunftsgleich oder identical by descent sind. Diese Wahrscheinlichkeit entspricht dem halben genetisch additiven Verwandtschaftsgrad  $a_{sd}$  zwischen den Tieren  $s$  und  $d$  und dem Inzuchtkoeffizienten  $F_i$  des Nachkommen  $i$  von  $s$  und  $d$ .

Bei der zweiten Interpretation liegt das **“infinitesimal model”** aus der quantitativen Genetik zu Grunde. In diesem Modell wird angenommen, dass Beobachtungswerte von quantitativen Eigenschaften von sehr vielen unabhängigen Genloci mit sehr kleiner Wirkung beeinflusst werden. Als weitere Annahme gehen wir im **“infinitesimal model”** auch davon aus, dass die genetischen Effekte unabhängig sind von der Umwelt und einer multivariaten Normalverteilung folgen. Durch die Selektion geht zwar die Eigenschaft der Normalverteilung verloren, aber (Bulmer, 1980) konnte zeigen, dass die Abweichungen nur von geringem Ausmass sind.

Basierend auf dem **“infinitesimal model”** und der Normalverteilung der Umwelteffekte und der phänotypischen Beobachtungen müssen auch die **Zuchtwerte** einer **multivariaten Normalverteilung** folgen. Diese Verteilung ist anhand der **Erwartungswerte** und der **Varianz-Kovarianz-Matrix** bestimmt. Schon beim Ein-Locus Modell wurden Zuchtwerte als Abweichungen definiert. Diese Definition führt dazu, dass der Erwartungswert von Zuchtwerten immer gleich null ist. Fassen wir alle Zuchtwerte einer Population in einen Vektor zusammen und bezeichnen den mit  $a$ , dann können wir einfach schreiben:

$$E[a] = 0$$

wobei 0 hier für den Null-Vektor steht.

Für die Varianz-Kovarianz-Matrix der Zuchtwerte hatten wir schon gesehen, dass wir diese gleich der genetisch-additiven Verwandtschaftsmatrix  $A$  mal der genetisch additiven Varianzkomponente  $\sigma_a^2$  gesetzt hatten. Somit gilt, dass

$$Var(a) = A * \sigma_a^2$$

### 9.6.1 Komponenten einer Varianz-Kovarianz-Matrix

Allgemein befinden sich in einer Varianz-Kovarianz-Matrix auf der Diagonalen die Varianzen der einzelnen Komponenten und auf der Off-Diagonalen die Kovarianzen zwischen den Komponenten.

### 9.6.2 Diagonalelemente

In unserem Falle handelt es sich bei den Komponenten um die Zuchtwerte aller Tiere in einer Population. Somit entspricht das  $i$ -te Diagonalelement  $((Var(a))_{ii})$  der Varianz-Kovarianz-Matrix  $Var(a)$  der Varianz des Zuchtwerte  $a_i$  von Tier  $i$ . Deshalb gilt die folgende Beziehung

$$(Var(a))_{ii} = Var(a_i) = (A)_{ii} * \sigma_a^2 = a_{ii} * \sigma_a^2 = (1 + F_i) * \sigma_a^2$$

Auf den ersten Blick scheint es schwer nachvollziehbar, dass mit zunehmendem Inzuchtgrad  $F_i$  die Varianz  $Var(a_i)$  auch zunehmen soll. Eine einfache Erklärung für dieses Phänomen folgt aus der Betrachtung des Ein-Locus-Modells. Wenn in einer Population der mittlere Inzuchtgrad steigt, dann steigt der Anteil Allele, welche herkunftsgleich sind. Wenn zwei Allele in einem Tier herkunftsgleich sind, dann ist das Tier sicher homozygot. Somit ist mit steigendem Inzuchtgrad in einer Population auch der Anteil der homozygoten Tiere erhöht. Im Gegenzug dazu sinkt der Anteil der heterozygoten Tiere. Für einen bestimmten Genort

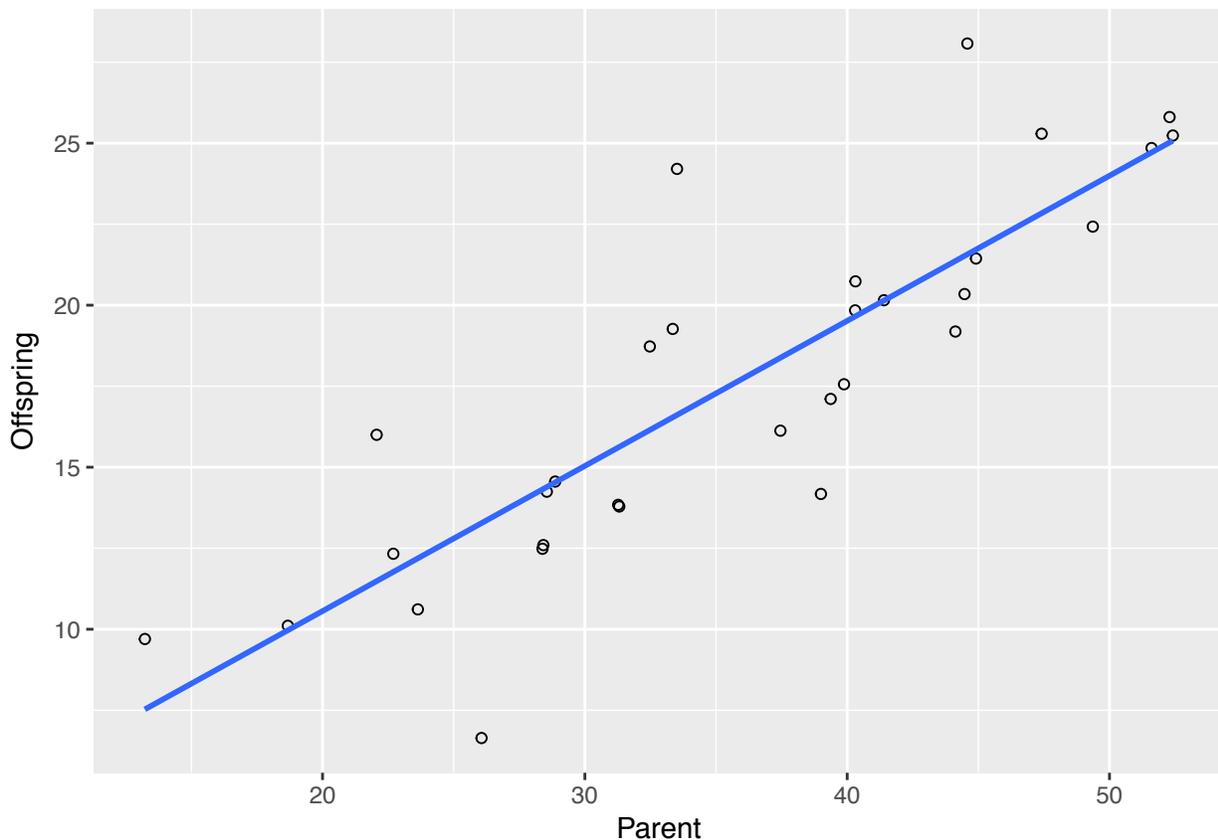
$G$  bedeutet dies, dass der Anteil an  $G_1G_2$  Tieren sinkt und der Anteil an  $G_2G_2$  und  $G_1G_1$  Tieren ansteigt. Im Bezug auf die Verteilung der genotypischen Werte in der Population nimmt der Anteil der extremen Werte von  $G_1G_1$  und  $G_2G_2$  zu im Vergleich zum Anteil der Werte von  $G_2G_2$ . Die Zunahme des Anteils der extremen Werte hat auch eine Zunahme der Varianz zur Folge.

### 9.6.3 Off-Diagonale

Auf der Off-Diagonalen stehen die Kovarianzen zwischen den Zuchtwerten verschiedener Tiere. Als Beispiel entspricht das Element der Matrix  $Var(a)$  in der dritten Zeile und der siebten Kolonne der Kovarianz der Zuchtwerte von Tier drei und Tier sieben. Dies entspricht gemäss unserer angenommenen Definition von  $Var(a)$  dem folgenden Ausdruck

$$(Var(a))_{37} = Cov(a_3, a_7) = (A)_{37} * \sigma_a^2 = a_{37} * \sigma_a^2$$

wobei  $a_{37}$  dem additiv genetischen Verwandtschaftsgrad zwischen den Tieren 3 und 7 entspricht. Hier stellt sich die Frage, weshalb wir die Kovarianz der Zuchtwerte zweier Tiere anhand des Verwandtschaftsgrades berechnen können. Nehmen wir einmal an, dass Tier 7 ein Nachkomme von Tier 3 ist. Die Frage ist nun, wie können wir die Kovarianz der Zuchtwerte für die Tiere 3 und 7 bestimmen? Stellen wir uns vor, dass wir für alle Eltern-Nachkommen-Paare in unserer Population eine phänotypische Leistung in einem Merkmal beobachten und das wir diese in einem Scatter-Plot auftragen. So ein Plot könnte etwa wie folgt aussehen:



Die Steigung  $b$  der in blau eingezeichneten Regressionsgeraden berechnet sich als Verhältnis zwischen der Kovarianz der Nachkommen- und Elternleistung und der Varianz der Elternleistung. Somit ist

$$b = \frac{Cov(y_{Offspring}, y_{Parent})}{Var(y_{Parent})}$$

Der Nenner in der obigen Formel des Regressionskoeffizienten entspricht einfach der phänotypischen Varianz der Elternleistungen und ist im Folgenden von keiner grossen Bedeutung. Für die Aufschlüsselung der Kovarianz im Zähler verwenden wir wieder unser Modell, welches phänotypische Leistungen in genetische Komponenten und in Umweltkomponenten aufteilt. Als genetische Komponenten interessieren uns hier nur die additiven Anteile (Zuchtwerte). Alle anderen Komponenten werden in einem zufälligen Restterm  $\epsilon$  zusammengefasst. Die phänotypischen Leistungen der Eltern lassen sich also aufteilen in

$$y_{Parent} = \mu + a_{Parent} + \epsilon_{Parent}$$

Analog dazu lassen sich die Leistungen der Nachkommen aufteilen in

$$y_{Offspring} = \mu + a_{Offspring} + \epsilon_{Offspring}$$

Unter der Annahme, dass die Eltern- und die Nachkommenleistungen in der gleichen Umwelt erbracht wurden und, dass sich die Zuchtwerte der anderen Elternteile im Mittel über alle Eltern-Nachkommen-Paare zu 0 aufaddieren, können wir den Zuchtwert der Nachkommen schreiben als

$$a_{Offspring} = 1/2 a_{Parent}$$

Setzen wir dies in der Kovarianz ein ergibt sich

$$\begin{aligned} Cov(y_{Offspring}, y_{Parent}) &= Cov(\mu + a_{Offspring} + \epsilon_{Offspring}, \mu + a_{Parent} + \epsilon_{Parent}) \\ &= Cov(a_{Offspring}, a_{Parent}) = Cov(1/2 a_{Parent}, a_{Parent}) = 1/2 * \sigma_a^2 \end{aligned}$$

Der Faktor vor  $\sigma_a^2$  entspricht genau dem Term in der genetisch-additiven Verwandtschaftsmatrix zwischen einem Elternteil und einem direkten Nachkommen. Mit der gleich Argumentation können auch alle anderen Koeffizienten von Verwandtenpaaren hergeleitet werden und wir werden in jedem Fall eine Übereinstimmung zwischen den Faktoren der Kovarianzen und den Elementen in der Verwandtschaftsmatrix  $A$  finden.

# Bibliography

Bulmer, M. (1980). *The mathematical theory of quantitative genetics*. Clarendon Press.

Dempfle, L. (1987). Problems in the use of the relationship matrix in animal breeding. International Symposium, Armidale. Notes for Participants.

Hofer, A. (1990). *Schätzung von Zuchtwerten feldgeprüfter Schweine mit einem Mehrmerkmals-Tiermodell*. PhD thesis, ETH Zürich.