



# Selektionsindex - Zuchtwert - Verwandtschaft

Peter von Rohr

# Zusammenfassung Zuchtwert aus Quantitativer Genetik

## Modell der phänotypischen Beobachtung

- Zerlegung der gemessenen phänotypischen Wertes (Beobachtung) ( $p$ )
- Komponenten sind *genotypischer Wert* ( $g$ ) und *Umweltabweichung* ( $e$ )
- $p = g + e$
- Erwartungswerte: Im Schnitt über eine Population werden die Umweltabweichungen  $e$  als Null angenommen
- $E[e] = 0$  und somit  $E[p] = E[g]$

# Genotypischer Wert (GW)

- GW erfasst den genetisch bedingten Teil des phänotypischen Wertes
- Annahme: 1 Genort, 2 Allele, Population im Hardy-Weinberg Gleichgewicht
- Für bestimmten Genotypen  $G_i G_j$  ist der genotypische Wert  $V_{ij}$  definiert als der mittlere Wert aller Individuen in der gleichen Umwelt mit Genotyp  $G_i G_j$



# Zusammenfassung Genotypische Werte

Genotyp	genotypischer Wert
$G_1 G_1$	$V_{11} = a$
$G_1 G_2$	$V_{12} = d$
$G_2 G_2$	$V_{22} = -a$

# Populationsmittel - Erwartungswert

- Populationsmittel  $\mu$  als Erwartungswert  $E[V]$  der genotypischen Werte
- Allgemeine Definition des Erwartungswertes einer diskreten Zufallsvariablen  $X$

$$E[X] = \sum_{x_i \in \mathcal{X}} x_i * f(x_i)$$

wobei  $\mathcal{X}$ : Menge aller möglichen  $x$ -Werte  
 $f(x_i)$  Wahrscheinlichkeit dass  $x$  den Wert  $x_i$  annimmt

# Erwartungswert der genotypischen Werte

Für unser Beispiel mit den genotypischen Werten

- Menge an möglichen Werten  $\mathcal{X} = \{V_{11}, V_{12}, V_{22}\}$
- Frequenzen der genotypischen Werte entspricht Frequenz der Genotypen
- Mit Allelfrequenzen  $f(G_1) = p$  und  $f(G_2) = q = 1 - p$  folgen die Genotypfrequenzen

Genotyp	Frequenzen
$G_1 G_1$	$p * p = p^2$
$G_1 G_2$	$p * q + q * p = 2pq$
$G_2 G_2$	$q * q = q^2$

# Populationsmittel als Erwartungswert

- Einsetzen der genotypischen Werte und der Genotypfrequenzen in Definition des Erwartungswertes

$$\begin{aligned}\mu &= E[V] \\ &= f(G_1 G_1) * V_{11} + f(G_1 G_2) * V_{12} + f(G_2 G_2) * V_{22}\end{aligned}\tag{1}$$

# Populationsmittel unter Hardy-Weinberg

Aufgrund der Genotypfrequenzen folgt

$$\begin{aligned}\mu &= p^2 * a + 2pq * d - q^2 * a && (2) \\ &= (p^2 - q^2)a + 2pqd \\ &= (p + q)(p - q)a + 2pqd \\ &= (p - q)a + 2pqd\end{aligned}$$

→ Das Populationsmittel ist somit abhängig von den Allelfrequenzen und somit von der Selektion

# Zuchtwert

- Bei Zucht interessiert, welche genetischen Anlagen von Eltern an Nachkommen weitergegeben werden
- Eltern geben nicht Genotypen sondern zufällige Stichprobe der Allele an Nachkommen weiter
- → Frage, welche Leistung kann von Nachkommen eines bestimmten Elterngenotyps erwartet werden
- Zuchtwert wird verwendet als Mass für die mittlere Leistung von Nachkommen eines bestimmten Tieres im Vergleich zum Populationsmittel
- Definition **Zuchtwert**: entspricht der doppelten Abweichung des erwarteten Mittelwertes von Nachkommen vom Populationsmittel

# Herleitung des Zuchtwertes für Genotyp $G_1 G_1$

- Frequenz der Nachkommen

	Vater	
	$f(G_1) = p$	$f(G_2) = q$
Mutter		
$f(G_1) = 1$	$f(G_1 G_1) = p$	$f(G_1 G_2) = q$

- Erwarteter mittlerer genotypischer Wert  $\mu_{11}$  der Nachkommen einer Mutter mit Genotyp  $G_1 G_1$

$$\mu_{11} = pa + qd$$

- Gemäss Definition, entspricht der Zuchtwert  $ZW_{11}$  der Mutter mit Genotyp  $G_1 G_1$  der doppelten Abweichung des mittleren genotypischen Wertes  $\mu_{11}$  der Nachkommen vom Populationsmittel  $\mu$

## Zuchtwert $ZW_{11}$ für Genotyp $G_1 G_1$

- Doppelte Abweichung der Nachkommen vom Populationsmittel
- Einsetzen der berechneten Größen für  $\mu_{11}$  und  $\mu$

$$\begin{aligned}
 ZW_{11} &= 2 * (\mu_{11} - \mu) && (3) \\
 &= 2(pa + qd - [(p - q)a + 2pqd]) \\
 &= 2(pa + qd - (p - q)a - 2pqd) \\
 &= 2(qd + qa - 2pqd) \\
 &= 2(qa + qd(1 - 2p)) \\
 &= 2q(a + d(1 - 2p)) \\
 &= 2q(a + (q - p)d)
 \end{aligned}$$

Falls Locus  $G$  rein additiv  $\rightarrow d = 0$ , somit ist  $ZW_{11} = 2qa$

# Zuchtwert $ZW_{22}$ für Genotyp $G_2G_2$

- Frequenz der Nachkommen

	Vater	
	$f(G_1) = p$	$f(G_2) = q$
Mutter		
$f(G_2) = 1$		

- Erwarteter mittlerer genotypischer Wert  $\mu_{22} =$
- Zuchtwert  $ZW_{22} =$

# Zuchtwert $ZW_{12}$ für Genotyp $G_1 G_2$

- Frequenz der Nachkommen

	Vater	
	$f(G_1) = p$	$f(G_2) = q$
Mutter		
$f(G_1) = 0.5$		
$f(G_2) = 0.5$		

- Erwarteter mittlerer genotypischer Wert  $\mu_{12} =$
- Zuchtwert  $ZW_{12} =$

# Allelsubstitution

- Bei allen Zuchtwerten kommt der Term  $a + (q - p)d$  vor. Dieser wird mit  $\alpha$  bezeichnet
- Zuchtwerte als Funktion von  $\alpha$

Genotyp	Zuchtwert
$G_1 G_1$	$2q\alpha$
$G_1 G_2$	$(q - p)\alpha$
$G_2 G_2$	$-2p\alpha$

## Allelsubstitution II

- Vergleicht man einen  $G_2G_2$  Genotyp mit einem  $G_1G_2$  Genotyp, dann wurde ein  $G_2$ -Allel durch ein  $G_1$ -Allel ersetzt
- Die Zuchtwerte ändern sich von  $ZW_{22}$  zu  $ZW_{12}$
- Änderung der Zuchtwerte:

$$\begin{aligned}ZW_{12} - ZW_{22} &= (q - p)\alpha - (-2p\alpha) && (4) \\ &= (q - p)\alpha + 2p\alpha \\ &= (q - p + 2p)\alpha \\ &= (q + p)\alpha \\ &= \alpha\end{aligned}$$

## Allelsubstitution III

- Analoger Vergleich zwischen Genotypen  $G_1G_2$  und  $G_1G_1$  führt zu

$$\begin{aligned} ZW_{11} - ZW_{12} &= 2q\alpha - (q - p)\alpha && (5) \\ &= (2q - (q - p))\alpha \\ &= \alpha \end{aligned}$$

- Zuchtwerte sind von den Allelfrequenzen abhängig
- Differenzen zwischen Zuchtwerten sind additiv

## Dominanzabweichung

- Zuchtwert definiert als doppelte Abweichung des erwarteten mittleren Wertes der Nachkommen vom Populationsmittel
- Somit wird für einen bestimmten Genotypen  $G_i G_j$  der genotypische Wert  $V_{ij}$  vom Zuchtwert  $ZW_{ij}$  abweichen.
- Für die Genotypen  $G_1 G_1$ ,  $G_1 G_2$  und  $G_2 G_2$  sehen die Abweichungen wie folgt aus

$$\begin{aligned}
 V_{11} - ZW_{11} &= a - 2q\alpha & (6) \\
 &= a - 2q[a + (q - p)d] \\
 &= a - 2qa - 2q(q - p)d \\
 &= a(1 - 2q) - 2q^2d + 2pqd \\
 &= [(p - q)a + 2pqd] - 2q^2d \\
 &= \mu + D_{11}
 \end{aligned}$$

## Dominanzabweichung II

- Für Genotyp  $G_1G_2$

$$\begin{aligned}
 V_{12} - ZW_{12} &= d - (q - p)\alpha & (7) \\
 &= d - (q - p)[a + (q - p)d] \\
 &= [(p - q)a + 2pqd] + 2pqd \\
 &= \mu + D_{12}
 \end{aligned}$$

- Für Genotyp  $G_2G_2$

$$\begin{aligned}
 V_{22} - ZW_{22} &= -a - (-2p\alpha) & (8) \\
 &= -a + 2p[a + (q - p)d] \\
 &= [(p - q)a + 2pqd] - 2p^2d \\
 &= \mu + D_{22}
 \end{aligned}$$

## Zusammenfassung Dominanzabweichung

Genotyp	genotypischer Wert	Zuchtwert	Dominanzabweichung
$G_i G_j$	$V_{ij}$	$ZW_{ij}$	$D_{ij}$
$G_1 G_1$	$a$	$2q\alpha$	$-2q^2 d$
$G_1 G_2$	$d$	$(q - p)\alpha$	$2pqd$
$G_2 G_2$	$-a$	$-2p\alpha$	$-2p^2 d$

- Die genotypischen Werte können mit folgendem Modell beschrieben werden

$$V_{ij} = \mu + ZW_{ij} + D_{ij}$$

# Varianzen

- Populationsmittel der genotypischen Werte ( $\mu = E(V)$ ) gibt Informationen zur Lage einer Population
- Varianz  $\sigma_G^2 = \text{Var}[V]$  gibt an, wie gross Streuung um Populationsmittel
- Definition der Varianz für diskrete Zufallsvariable  $X$

$$\text{Var}[X] = \sum_{x_i \in \mathcal{X}} (x_i - \mu_X)^2 * f(x_i)$$

wobei  $\mathcal{X}$ : Menge aller möglichen  $x$ -Werte  
 $f(x_i)$ : Wahrscheinlichkeit dass  $x$  den Wert  $x_i$  annimmt  
 $\mu_X$ : Erwartungswert  $E[X]$  von  $X$

## Varianzen der genotypischen Werte

Einsetzen der Frequenzen und der genotypischen Werte in die Definition der Varianz

$$\begin{aligned}
 \sigma_G^2 = \text{Var}[V] &= (V_{11} - \mu)^2 * f(G_1 G_1) && (9) \\
 &+ (V_{12} - \mu)^2 * f(G_1 G_2) \\
 &+ (V_{22} - \mu)^2 * f(G_2 G_2) \\
 &= (a - \mu)^2 * p^2 \\
 &+ (d - \mu)^2 * 2pq \\
 &+ (-a - \mu)^2 * q^2
 \end{aligned}$$

wobei  $\mu = (p - q)a + 2pqd$  das Populationsmittel

## Herleitung der Varianz

Zur Herleitung der Varianz  $\sigma_G^2$  verwenden wir, dass  $V_{ij} = \mu + ZW_{ij} + D_{ij}$  und somit

$$\begin{aligned}
 \sigma_G^2 &= (ZW_{11} + D_{11})^2 * p^2 && (10) \\
 &+ (ZW_{12} + D_{12})^2 * 2pq \\
 &+ (ZW_{22} + D_{22})^2 * q^2 \\
 &= (2q\alpha - 2q^2d)^2 * p^2 \\
 &+ ((q - p)\alpha + 2pqd)^2 * 2pq \\
 &+ (-2p\alpha - 2p^2d)^2 * q^2 \\
 &\dots \\
 &= 2pq\alpha^2 + (2pqd)^2
 \end{aligned}$$

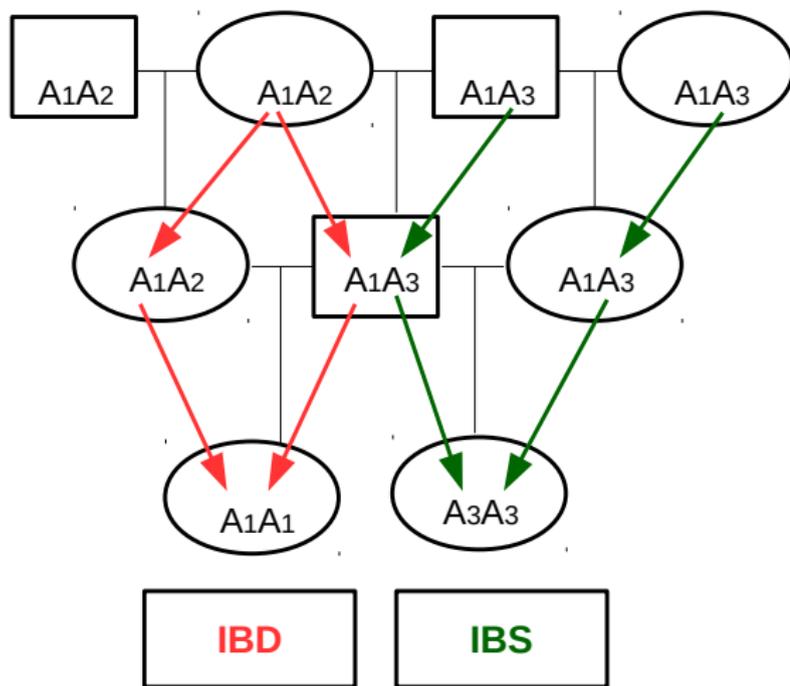
# Aufteilung der Varianz

- Aufteilung von  $\sigma_G^2 = 2pq\alpha^2 + (2pqd)^2$  in die zwei Terme
- Term  $2pq\alpha^2$  heisst **genetisch additive Varianz** und wird mit  $\sigma_A^2$  bezeichnet
- Term  $(2pqd)^2$  heisst **Dominanzvarianz** und wird mit  $\sigma_D^2$  bezeichnet.
- Somit ist  $\sigma_G^2 = \sigma_A^2 + \sigma_D^2$

# Verwandtschaft

- Bei künstlicher Selektion sollen Tiere mit günstigen Eigenschaften ausgewählt werden
- Ausgewählte Elterntiere sind in wichtigen Eigenschaften ähnlich
- Herkunft der Ähnlichkeit ist entscheidend
- Es wird unterschieden zwischen **zustandsgleich** (identical by state - IBS) und **herkunftsgleich** (identical by descent - IBD)

## IBD vs. IBS



## Definition Verwandtschaft

- Zwei Tiere  $x$  und  $y$  sind verwandt, wenn die Wahrscheinlichkeit, dass sie an einem beliebigen Locus herkunftsgleiche Allele (IBD) aufweisen,  $> 0$  ist.
- **Verwandtschaftsgrad**  $a_{xy}$  beschreibt die mittlere Wahrscheinlichkeit über alle Loci, dass diese IBD sind
- Allgemeine Berechnung des Verwandtschaftsgrades  $a_{xy}$  zwischen Tieren  $x$  und  $y$

$$a_{xy} = \sum_{\text{Pfade}} \left(\frac{1}{2}\right)^{t_1+t_2} (1 + F_Z)$$

wobei  $\sum_{\text{Pfade}}$  die Summe über alle Pfade im Pedigree meint,  $t_1$  und  $t_2$  die Anzahl Generationen von  $x$  und  $y$  zum gemeinsamen Ahnen  $Z$  und  $F_Z$  den Inzuchtgrad von  $Z$

# Inzucht

- Tier  $x$  ist ingezüchtet, falls seine Eltern  $M_x$  und  $V_x$  verwandt sind miteinander
- Inzucht beschreibt, ob in einem Tier  $x$  an einem beliebigen Locus herkunftsgleiche Allele vorliegen
- Sind Allele IBD, dann müssen sie Kopien vom gleichen Ahnenallel sein.
- Berechnung des Inzuchtgrades für das Tier  $x$

$$F_x = \frac{1}{2} a_{M_x, V_x} = \frac{1}{2} \sum_{\text{Pfade}} \left(\frac{1}{2}\right)^{t_1+t_2} (1+F_Z) = \sum_{\text{Pfade}} \left(\frac{1}{2}\right)^{t_1+t_2+1} (1+F_Z)$$

# Matrixmethode

Zur Berechnung der Verwandtschaftsgrade und der Inzuchtgrade wurde folgende Methode entwickelt

- Tiere dem Alter nach von links nach rechts in einer symmetrischen Matrix anordnen
- Oberhalb jedes Tieres werden Eltern angefügt
- Ausfüllen der Matrix getrennt nach Diagonalelementen und anderen
  - 1 Diagonalelemente:  $1 + F_x$
  - 2 Off-Diagonal Elemente:  $a_{xy} = \frac{1}{2}(a_{x,M_y} + a_{x,V_y})$

# Beispielpedigree

