

Population structure and genetic effects

Prof. Roberto Fritsche-Neto

roberto.neto@usp.br

Piracicaba, August 21th, 2019

Non-inbred populations

- Population = a pool of shared alleles
- Only the allele is passed to the offspring, not the genotype
- **Allele frequencies**
- $B = f(B) = p$
- $= p^2 + \frac{1}{2}(2pq)$
- $= p^2 + pq$
- $= p^2 + p(1-p)$
- $= p^2 + p - p^2$
- $= p$
- $b = f(b) = q$
- $= q^2 + \frac{1}{2} 2pq$
- $= q^2 + pq$
- $= q^2 + q(1-q)$
- $= q^2 + q - q^2$
- $= q$
- **Under HWE these frequencies are kept constant**
- **Otherwise, with just one random mating this equilibrium is reached again**

Genotype	f
BB	p^2
Bb	$2pq$
bb	q^2

Inbred populations

- **Inbred populations**
- Alleles identical by descendent (**IBD**) – copies of a shared ancestral
- Alleles identical by state (**IBS**) – same allele, but different origin

- **Inbreeding coefficient**

- $F = P(B_i \equiv B_j)$
- $F = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

Allele	B	b
B	BB (1/4) IBD	Bb (1/4)
b	bB (1/4)	bb (1/4) IBD

- **Probability of IBD**

- It is equal of the allele frequency times the inbreeding rate
- $P(BB) = Fp$
- $P(bb) = Fq$

- **Non-inbred genotypes** = $1 - F$

- **Inbred genotypes** = F

Inbred populations

- **IBD** **IBS**
- $= F(BB + bb) + (1 - F)(BB + Bb + bb)$
- $= F(pBB + qbb) + (1 - F)(p^2BB + 2pqBb + q^2bb)$
- $= FpBB + Fqbb + p^2BB + 2pqBb + q^2bb - Fp^2BB - 2FpqBb - Fq^2bb$
- $= BB[Fp + p^2 - Fp^2] + Bb[2pq - 2Fpq] + bb[Fq + q^2 - Fq^2]$
- $= BB[p^2 + Fp - Fp(1-q)] + Bb[2pq - 2pqF] + bb[q^2 + Fq - Fq(1-p)]$
- **Then, for any F**
- $= BB[p^2 + pqF] + Bb[2pq - 2pqF] + bb[q^2 + pqF]$

- **F = 1**
- $p^2 + pqF$
- $p^2 + pq$
- $p^2 + p(1-p)$
- $p^2 + p - p^2 = p$

Genotype	f	$F=0$	$F=1$
BB	$p^2 + pqF$	p^2	p
Bb	$2pq - 2pqF$	$2pq$	0
bb	$q^2 + pqF$	q^2	q

Genotypes and genic effects

- **Intra population** - assuming HWE
- α = female i and male j allele effects at the k locus
- $\mathcal{S}$ = interaction between alleles i and j
- The sum of the α over k loci means *Breeding value (BV)*

- **Substitution allele effect**

- $\alpha_B = u_B - u_{\text{pop}}$

- $\alpha_b = u_b - u_{\text{pop}}$

- **Crosses**

- BB x population = u_B

- bb x population = u_b

- $u_B = pa + qd$

- $u_b = q(-a) + pd$

- $u_{\text{pop}} = p^2a + q^2(-a) + 2pqd$

- $= (p^2 - q^2)a + 2pqd$

- $= (p - q)a + 2pqd$

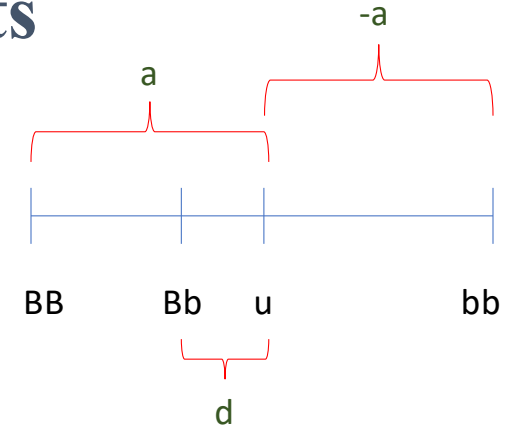
- $(p^2 - q^2) = (1 - q)^2 - q^2 = 1 - 2q + q^2 - q^2 = p + q - 2q = p - q$

$$y_i = u + g_i + g_j + s_{ij}$$

$$G_{ij} = u + \alpha_i + \alpha_j + \delta_{ij}$$

Genotype	f	VG
BB	p^2	a
Bb	$2pq$	d
bb	q^2	$-a$

Genotype	f	Mean	Dominance
BB	p^2	$\alpha_B + \alpha_B$	$\mathcal{S}_{BB}$
Bb	$2pq$	$\alpha_B + \alpha_b$	$\mathcal{S}_{Bb}$
bb	q^2	$\alpha_b + \alpha_b$	$\mathcal{S}_{bb}$



Substitution allele effect

- $\alpha_B = u_B - u_{\text{pop}}$
- $= pa + qd - [(p - q)a + 2pqd]$
- $= pa + qd - pa + qa - 2pqd$
- $= qa + qd - 2pqd$
- $= q[a + (1 - 2p)d]$
- $= q[a + (q + p - 2p)d]$
- $= q[a + (q - p)d]$
- $= q\alpha$

- $\alpha_b = u_b - u_{\text{pop}}$
- $= -qa + pd - [(p - q)a + 2pqd]$
- $= -qa + pd - pa + qa - 2pqd$
- $= pd - pa - 2pqd$
- $= -p[a - d + 2qd]$
- $= -p[a(2q - 1)d]$
- $= -p[a(2q - p + q)d]$
- $= -p[a + (q - p)d]$
- $= -p\alpha$

Dominance effect

- $BB = u + 2\alpha_B + \mathcal{S}_{BB} = a$
- $\mathcal{S}_{BB} = a - u - 2\alpha_B$
- $= a - [(p - q)a + 2pqd] - 2[q[a + (q - p)d]$
- $= a - pa + qa - 2pqd - 2qa - 2q(q - p)d$
- $= a - pa - qa - 2pqd - 2q^2d + 2pqd$
- $= a - pa - qa - 2q^2d$
- $= a(1 - p - q) - 2q^2d$
- $= a(p + q - p - p) - 2q^2d$
- $= -2q^2d$
- $bb = u + 2\alpha_b + \mathcal{S}_{bb} = -a$
- $\mathcal{S}_{bb} = -a - u - 2\alpha_b$
- $= -a - [(p - q)a + 2pqd] - 2[-p[a + (q - p)d]$
- $= -a - pa + qa - 2pqd + 2pa + 2p(q - p)d$
- $= -a + pa + qa - 2pqd - 2p^2d + 2pqd$
- $= -a + pa + qa - 2p^2d$
- $= a(p + q - 1) - 2p^2d$
- $= a(p + q - p - q) - 2p^2d$
- $= -2p^2d$

- $Bb = u + \alpha_B + \alpha_b + \mathcal{S}_{Bb} = d$
- $\mathcal{S}_{Bb} = d - u - \alpha_B - \alpha_b$
- $= d - [(p - q)a + 2pqd] - [q[a + (q - p)d] - [-p[a + (q - p)d]$
- $= d - pa + qa + 2pqd - qa - q(q - p)d + pa - p(q - p)d$
- $= d - 2pqd - q(q - p)d - p(q - p)d$
- $= d - 2pqd - q^2d + pqd - p^2d + pqd$
- $= d - q^2d - p^2d$
- $= d(1 - q^2 - p^2)$
- $= d(1 - (1 - p)^2 - p^2)$
- $= d(1 - (1 + p^2 + 2p) - p^2)$
- $= d(1 - 1 - p^2 - 2p - p^2)$
- $= d(2p - 2p^2)$
- $= 2d[p - p(1 - q)]$
- $= 2d[p - p + pq]$
- $= 2d[pq]$
- $= 2pqd$

Genotype	f	Mean	Dominance
BB	p^2	$\alpha_B + \alpha_B$	$\mathcal{S}_{BB}$
Bb	$2pq$	$\alpha_B + \alpha_b$	$\mathcal{S}_{Bb}$
bb	q^2	$\alpha_b + \alpha_b$	$\mathcal{S}_{bb}$

Substitution allele effect

- α is the slope of the regression line
- Regression coefficient – b

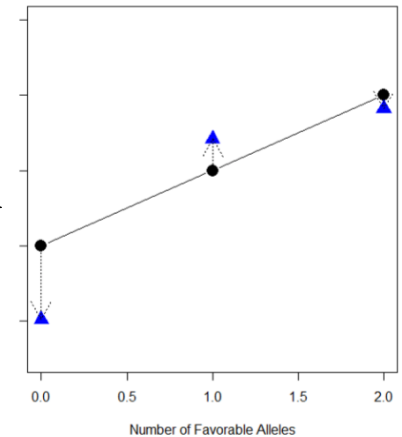
$$b = \frac{COV(x,y)}{V_x}$$

$$\sigma^2 = \sum_i f_i X_i^2 - \left(\sum_i f_i X_i \right)^2 \quad COV_{xy} = \sum_i f_i X_i Y_i - \left(\sum_i f_i X_i \cdot \sum_i f_i Y_i \right)$$

Genotype	f	Alleles x	VG y
BB	p^2	2	a
Bb	$2pq$	1	d
bb	q^2	0	$-a$

- Thus, the variance of x is given by:
- $p^2(2^2) + 2pq(1^2) + q^2(0^2) = 4p^2 + 2pq$
- $[p^2(2) + 2pq(1) + q^2(0)]^2 = [2p^2 + 2pq]^2 = [2p^2 + 2p(1 - p)]^2 = [2p^2 + 2p - 2p^2]^2 = [2p]^2 = 4p^2$
- $V_x = 4p^2 + 2pq - 4p^2 = 2pq$
- and the covariance by
- $p^2(a)(2) + 2pq(d)(1) + q^2(-a)(0) = 2p^2a + 2pqd$
- $p^2(2) + 2pq(1) + q^2(0) = 2p^2 + 2pq = 2p$
- $p^2(a) + 2pq(d) + q^2(-a) = p^2a + 2pqd - q^2a = (p^2 - q^2)a + 2pqd = (p - q)a + 2pqd$
- $2p(2p^2a + 2pqd) = 2p(p - q)a + 4p^2qd$
- $COV(x, y) = 2p^2a + 2pqd - 2p(p - q)a + 4p^2qd = 2pq[a + (q - p)d]$

$$b = \frac{COV(x,y)}{V_x} = \frac{2pq[a + (q - p)d]}{2pq} = a + (q - p)d = \alpha$$



Additive variance

- $V(X) = E[X - E(X)]^2$

$$G_{ij} = u + \alpha_i + \alpha_j + \delta_{ij}$$

- $Vg = E[G_{ij} - E(G_{ij})]^2$
- $= E[u + \alpha_i + \alpha_j + \delta_{ij} - u]^2$
- $= E[\alpha_i + \alpha_j + \delta_{ij}]^2$
- $= E[\alpha_i]^2 + E[\alpha_j]^2 + E[\delta_{ij}]^2 + \dots$
- $= E[\alpha_i]^2 + E[\alpha_j]^2 + E[\delta_{ij}]^2$

$$\sigma_A^2 = 2 \sum_i p_i (\alpha_i)^2$$

- **Additive variance** = variance of the average allele effects

- $= E[\alpha_i]^2 = E[p_i(\alpha_i)^2 - E(p_i \alpha_i)]^2$
- $= E[p_i(\alpha_i)^2 - 0]^2$
- $= \sum p_i(\alpha_i)^2 = \frac{1}{2} Va$

- $= E[\alpha_j]^2 = E[p_j(\alpha_j)^2 - E(p_j \alpha_j)]^2$
- $= E[p_j(\alpha_j)^2 - 0]^2$
- $= \sum p_j(\alpha_j)^2 = \frac{1}{2} Va$

- $Va = 2 \cdot p \cdot \alpha_B^2 + 2 \cdot q \cdot \alpha_b^2$
- $= 2 \cdot p \cdot (q\alpha)^2 + 2 \cdot q \cdot (-p\alpha)^2$
- $= 2pq^2\alpha^2 + 2qp^2\alpha^2$
- $= 2pq\alpha^2(q + p)$
- $= 2pq\alpha^2$
- $= 2pq[a + (q - p)d]^2$

- $A = \alpha_i + \alpha_j = E(\alpha_B + \alpha_b) = 0$

- $E(A) = E(\alpha_i + \alpha_j) = E(\alpha_i) + E(\alpha_j)$
- $p_i \cdot \alpha_i + p_j \alpha_j$
- $p_i(\alpha_B) + p_j(\alpha_b)$
- $p(q\alpha) + q(-p\alpha)$
- $p(q\alpha) + q(-p\alpha)$
- $pq\alpha - pq\alpha = 0$

Dominance

- **Dominance variance** = interaction between allele effects within a locus

$$\begin{aligned}
 &= E[\mathcal{S}_{ij}]^2 = E[p_i(\mathcal{S}_{ij})^2 - E(p_i\mathcal{S}_{ij})]^2 \\
 &= E[p_i(\mathcal{S}_{ij})^2 - 0]^2 \\
 &= p_i(\mathcal{S}_{ij})^2 = \mathbf{Vd}
 \end{aligned}$$

$$G_{ij} = u + \alpha_i + \alpha_j + \delta_{ij}$$

$$\sigma_D^2 = \sum_i p_i p_j (\mathcal{S}_{ij})^2$$

Genotype	f	value	Dominance
BB	p^2	$-2q^2d$	$\mathcal{S}_{BB}$
Bb	$2pq$	$2pqd$	$\mathcal{S}_{Bb}$
bb	q^2	$-2p^2d$	$\mathcal{S}_{bb}$

$$\begin{aligned}
 &Vd = p^2 \cdot \mathcal{S}_{BB}^2 + 2pq \cdot \mathcal{S}_{Bb}^2 + q^2 \cdot \mathcal{S}_{bb}^2 \\
 &= p^2(-2q^2d)^2 + 2pq(2pqd)^2 + q^2(-2p^2d)^2 \\
 &= p^2(4q^4d^2) + 2pq(4p^2q^2d^2) + q^2(4p^4d^2) \\
 &= 4p^2q^2d^2(q^2 + 2pq + p^2) \\
 &= 4p^2q^2d^2(p + q)^2 \\
 &= (2pqd)^2
 \end{aligned}$$

$$\begin{aligned}
 &E(D) = \sum_{ij} (p_i p_j \mathcal{S}_{ij}) \\
 &= p^2 \cdot \mathcal{S}_{BB} + 2pq \cdot \mathcal{S}_{Bb} + q^2 \cdot \mathcal{S}_{bb} \\
 &= p^2(-2q^2d) + 2pq(2pqd) + q^2(-2p^2d) \\
 &= -2p^2q^2d + 4p^2q^2d + 2p^2q^2d \\
 &= 0
 \end{aligned}$$

Mean in inbred populations

- **IBD** **IBS**
- $= F(BB + bb) + (1-F)(BB + Bb + bb)$
- $= BB[p^2 + pqF] + Bb[2pq - 2pqF] + bb[q^2 + pqF]$

$$G_{ij} = u + \alpha_i + \alpha_j + \delta_{ij}$$

Genotype	f	$F=0$	$F=1$
BB	$p^2 + pqF$	p^2	p
Bb	$2pq - 2pqF$	$2pq$	0
bb	$q^2 + pqF$	q^2	q

- $E(G_{ij}) = F \sum_i p_i (u + \alpha_i + \alpha_i + S_{ii}) + (1-F) \sum_{ij} p_i p_j (u + \alpha_i + \alpha_j + S_{ij})$
- $= Fu + 2F \sum_i p_i \alpha_i + F \sum_i p_i S_{ii} + (1-F)u + (1-F) \sum_i p_i \alpha_i + (1-F) \sum_j p_j \alpha_j + (1-F) \sum_{ij} p_i p_j S_{ij}$
- $= Fu + (1-F)u + (1-F) \sum_i p_i S_{ii}$
- $= u + F \sum_i p_i S_{ii}$
- **Mean + inbreeding depression (ID)**
- $F \sum_i p_i S_{ii} = Fp S_{BB} + Fq S_{bb}$
- $= Fp(-2q^2d) + Fq(-2p^2d)$
- $= F(-2pq^2d) + F(-2qp^2d)$
- $= -2pqdF$
- **Thus,**
- $u_F = u - 2pqdF$
- $u_F = (p - q)a + 2pqd - 2pqdF$

Variance in inbred populations

- $Vg = E[G_{ij} - E(G_{ij})]^2$
- $= E[u + \alpha_i + \alpha_j + S_{ij} - u - F\sum_i p_i S_{ii}]^2$
- $= E[\alpha_i + \alpha_j + S_{ij} - F\sum_i p_i S_{ii}]^2$
- $= E[\alpha_i]^2 + E[\alpha_j]^2 + E[S_{ij}]^2 + E(-F\sum_i p_i S_{ii})^2 + dp$

- $= E[\alpha_i]^2 = E[\sum p_i (\alpha_i)^2 - E(p_i \alpha_i)]^2$
- $= E[p_i (\alpha_i)^2 - 0]^2$
- $= \sum p_i (\alpha_i)^2 = \frac{1}{2} Va$

- $= E[\alpha_j]^2 = E[p_i (\alpha_j)^2 - E(p_i \alpha_j)]^2$
- $= E[p_i (\alpha_j)^2 - 0]^2$
- $= \sum p_i (\alpha_j)^2 = \frac{1}{2} Va$

- $= E[S_{ij}]^2 = (1 - F)\sum_{ij} p_i p_j (S_{ij})^2 - F\sum_i p_i (S_{ii})^2$
- $\quad \quad \quad \text{non-inbred} \quad \quad \quad \text{inbred}$
- $= (1 - F)Vd - F\sum_i p_i (S_{ii})^2$

- $E(-F\sum_i p_i S_{ii})^2 = (F\sum_i p_i S_{ii})^2$

- $2E(\alpha_i \alpha_j) = 2(1 - F)E(\alpha_i)E(\alpha_j) + 2F\sum_i p_i (\alpha_i)^2$
- $= 0 + 2F \sum p_i (\alpha_i)^2$
- $= 2F \frac{1}{2} Va = FVa$

- $2E(\alpha_i S_{ij}) = 2(1 - F)E(\alpha_i)E(S_{ij}) + 2F \sum_i p_i \alpha_i S_{ii}$
- $= 0 + 2F \sum_i p_i \alpha_i S_{ii}$

- $2E(\alpha_i S_{jj}) = 2(1 - F)E(\alpha_i)E(S_{jj}) + 2F \sum_j p_j \alpha_j S_{jj}$
- $= 0 + 2F \sum_j p_j \alpha_j S_{jj}$

- $2E(\alpha_i \cdot - F\sum_i p_i S_{ii}) = 2E(\alpha_i)(-F\sum_i p_i S_{ii})$
- $= 0$

- $2E(\alpha_j \cdot - F\sum_j p_j S_{jj}) = 2E(\alpha_j)(-F\sum_j p_j S_{jj})$
- $= 0$

Variance in inbred populations

- $2E(\mathcal{S}_{ij} - F\sum_i p_i \mathcal{S}_{ii}) = (1 - F)E(\mathcal{S}_{ij})(-F\sum_i p_i \mathcal{S}_{ii}) + 2FE(\mathcal{S}_{ii})(-F\sum_i p_i \mathcal{S}_{ii})$
- $= 0 + (2F\sum_i p_i \mathcal{S}_{ii})(-F\sum_i p_i \mathcal{S}_{ii})$
- $= -2F^2(\sum_i p_i \mathcal{S}_{ii})^2$
- **Finally,**
- $V_g = \frac{1}{2} V_a + \frac{1}{2} V_a + (1 - F)V_d - F\sum_i p_i (\mathcal{S}_{ii})^2 + 2F\sum_i p_i \alpha_i \mathcal{S}_{ii} + 2F\sum_j p_j \alpha_j \mathcal{S}_{jj} - 2F^2(\sum_i p_i \mathcal{S}_{ii})^2$
- $V_g = (1+F)V_a + (1 - F)V_d + \underbrace{2F \sum_i p_i \alpha_i \mathcal{S}_{ii} + 2F \sum_j p_j \alpha_j \mathcal{S}_{jj}}_{D_1} + (F\sum_i p_i \alpha_i \mathcal{S}_{ii})^2 - 2(F^2\sum_i p_i \mathcal{S}_{ii}^2)$
- $D_1 = \frac{1}{2} (\sum_i p_i \alpha_i \mathcal{S}_{ii} + \sum_j p_j \alpha_j \mathcal{S}_{jj}) = \sum_i p_i \alpha_i \mathcal{S}_{ii}$
- Covariance between additive and dominance effects in the homozygotes
- $V_g = (1+F)V_a + (1 - F)V_d + 4FD_1 + F\sum_i p_i (\mathcal{S}_{ii})^2 - F^2(\sum_i p_i \mathcal{S}_{ii})^2$
- $V_g = (1+F)V_a + (1 - F)V_d + 4FD_1 + F\underbrace{[\sum_i p_i (\mathcal{S}_{ii})^2 - (\sum_i p_i \mathcal{S}_{ii})^2]}_{D_2} + F(1 - F)\underbrace{(\sum_i p_i \mathcal{S}_{ii})^2}_H$
- D_2 = Variance due to the dominance effects in the homozygotes
- H = the square of the inbreeding depression
- $V_g = (1+F)V_a + (1 - F)V_d + 4FD_1 + FD_2 + F(1 - F)H$

Variance at any level of inbreeding

- $V_g = (1 + F)V_a + (1 - F)V_d + 4FD_1 + FD_2 + F(1 - F)H$
- $D_1 = COV(a, d)$
- $D_2 = V(S_{ii})$
- $H = \text{the square of the inbreeding depression}$
- $F = 0$
- $V_g = (1 + 0)V_a + (1 - 0)V_d + 4 \cdot 0 \cdot D_1 + 0D_2 + 0(1 - 0)H$
- $V_g = V_a + V_d$
- $F = 1$
- $V_g = (1 + 1)V_a + (1 - 1)V_d + 4 \cdot 1 \cdot D_1 + 1D_2 + 1(1 - 1)H$
- $V_g = 2V_a + 4D_1 + D_2$
- **Inbred progenies** = Normally, D_1 is negative, reducing the V_g
- On the other hand, the bigger F , the bigger V_a

Mixed populations

- s = inbreeding rate
- $s = 2F / (1 + F)$
- $F = s / (2 - s)$
- $Vg = (1 + F)Va + (1 - F)Vd + 4FD_1 + FD_2 + F((1 - F)H)$

$$Vg = \left(\frac{2}{2-s}\right)Va + \left(\frac{2-2s}{2s}\right)Vd + \left(\frac{4s}{2s}\right)D1 + \left(\frac{s}{2s}\right)D2 + \left[\frac{2(1-s)}{2s^2}\right]H$$

- *Considering two alleles*
 - $Va = 2 \sum p_i(\alpha_i)^2 = 2pq\alpha^2$
 - $Vd = \sum_{ij} p_i p_j (\mathcal{S}_{ij})^2 = (2pqd)^2$
 - $D_1 = \sum_i p_i \alpha_i \mathcal{S}_{ii} = 2pq(q-p)(a + (q-p)d)d = 2pq(q-p)\alpha d$
 - $D_2 = \sum_i p_i (\mathcal{S}_{ii})^2 - (\sum_i p_i \mathcal{S}_{ii})^2 = 4pqd^2(q-p)^2$
 - $H = (\sum_i p_i \mathcal{S}_{ii})^2 = (2pqd)^2$
- *Parametric space*
 - $Va \geq 0$
 - $Vd \geq 0$
 - $-\infty \leq D_1 \leq +\infty$
 - $D_2 \geq 0$
 - $H \geq 0$

D_1 and D_2

- $D_1 = \sum_i p_i \alpha_i \mathcal{S}_{ii}$
- $= p \alpha_B \mathcal{S}_{BB} + q \alpha_b \mathcal{S}_{bb}$
- $= p(q\alpha)(-2q^2d) + q(-p\alpha)(-2p^2d)$
- $= 2pq^3\alpha d + 2p^3q\alpha d$
- $= -2pq(q-p)\alpha d$
- $D_2 = \sum_i p_i (\mathcal{S}_{ii})^2 - (\sum_i p_i \mathcal{S}_{ii})^2$
- $= [p(-2q^2d)^2 + q(-2p^2d)^2] - [p(-2q^2d) + q(-2p^2d)]^2$
- $= [4pq^4d^2 + 4qp^4d^2] - [-2pq^2d - 2qp^2d]^2$
- $= [4pqd^2(p^3 + q^3)] - [-2pqd(q + p)]^2$
- $= [4pqd^2(p^3 + q^3)] - 4p^2q^2d^2$
- $= 4pqd^2(p^3 + q^3 - pq)$
- $= 4pqd^2(p^3 + (1-p)^3 - p(1-p))$
- $= 4pqd^2(p^3 + (1-p)^3 - p + p^2)$
- $= 4pqd^2(1-2p)^2$
- $= 4pqd^2(q-p)^2$