

# Lecture 3

## Resemblance Between Relatives

Bruce Walsh. July 2005. Asian Institute on Statistical Genetics

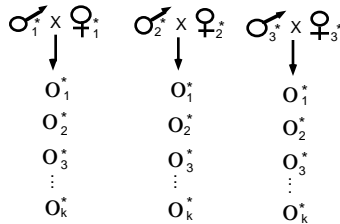
The **heritability** of a trait, a central concept in quantitative genetics, is the proportion of variation among individuals in a population that is due to variation in the additive genetic (i.e., breeding) values of individuals:

$$h^2 = \frac{V_A}{V_P} = \frac{\text{Variance of breeding values}}{\text{Phenotypic Variance}}$$

Since an individual's phenotype can be directly scored, the phenotypic variance  $V_P$  can be estimated from measurements made directly on the population.

In contrast, an individual's breeding value cannot be observed directly, but rather must be inferred from the mean value of its offspring (or more generally using the phenotypic values of other known relatives). Thus estimates of  $V_A$  require known collections of relatives. The most common situations (which we focus on here) are comparisons between parents and their offspring or comparisons among sibs. We can classify relatives as either **ancestral** or **collateral**, and we focus here on designs with just one type of relative. In a more general pedigree, information from both kinds of relatives is present.

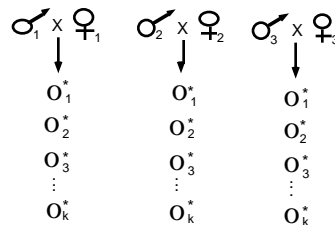
**Ancestral** relatives: e.g., parent and offspring



\* Measure phenotypes of one or both parents, and  $k$  offspring

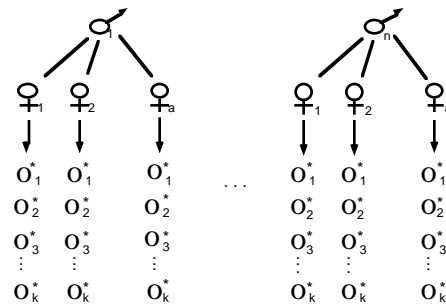
**Collateral** relatives:

Full Sibs have both parents in common



\*Measure  $k$  offspring in each family, but not the parents.

Half Sibs have one parent in common



\* Measure phenotype of  $k$  progeny of each family, but not the parents. Note that if  $k > 1$ , this design involves both full- (within any column) and half-sibs (between columns from the same sire) and is referred to as a **nested half-sib/full-sib design**.

**Key observation:** *The amount of phenotypic resemblance among relatives for the trait provides an indication of the amount of genetic variation for that trait. If trait variation has a significant genetic basis, the closer the relatives, the more similar their appearance.*

### Phenotypic Resemblance Between Relatives

Quantitative genetics as a field traces back to R. A. Fisher's 1918 paper showing how the phenotypic covariance between relatives can be expressed in terms of genetic variances, as we detail below.

#### 1. Parent-offspring regressions

There are three types of parent-offspring regressions: two **single parent - offspring regressions** (plotting offspring mean versus either the trait value in their father  $P_f$  or their mother  $P_m$ ), and the **midparent-offspring regression** (the offspring mean regressed on the mean of their parents, the **midparent**  $MP = (P_f + P_m)/2$ ).

The slope of the (single) parent-offspring regression is estimated by

$$b_{O|P} = \frac{Cov(O, P)}{Var(P)}, \quad \text{where} \quad Cov(O, P) = \frac{1}{n-1} \left( \sum_{i=1}^n O_i P_i - n \bar{O} \cdot \bar{P} \right)$$

where  $O_i$  is the mean trait value in the offspring of parent  $i$ ,  $\bar{O}$  the offspring mean over all families (also called the **grand mean** of all offspring),  $\bar{P}$  the mean of all parents, and we examine  $n$  parent-offspring families. One could compute separate regressions using males ( $P_m$ ) and females ( $P_f$ ), although the later potentially includes maternal effect contributions and hence single-parent regressions are usually restricted to fathers.

The midparent-offspring regression slope is estimated by

$$b_{O|MP} = \frac{Cov(O, MP)}{Var(MP)}, \quad \text{where} \quad Cov(O, MP) = \frac{1}{n-1} \left( \sum_{i=1}^n O_i P_{mp,i} - n \bar{O} \cdot \overline{MP} \right)$$

where  $O_i$  is the mean trait value in the offspring of parents in pair  $i$ , where these parents have an average trait value  $MP_i$  and we examine  $n$  parent-offspring pairs.

Notice that all of the three regressions involve the covariance between parents and their offspring.

## 2. Collateral relationships: ANOVA

With collateral relatives, the above formulae for the sample covariance is not appropriate, for two reasons. First, there are usually more than two collateral relatives per family. Second, even if families consist of only two relatives, the order of the two is arbitrary — i.e., there is no natural distinction between “X” and “Y”, as exists in the case of parents and offspring.

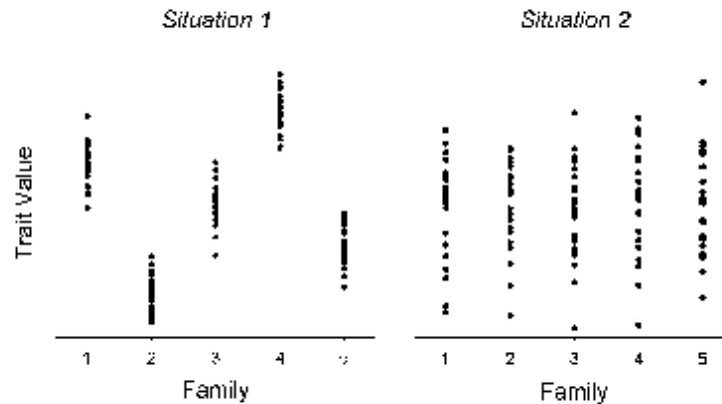
Another way of stating the second point is that collateral relatives belong to the same class or category. In contrast, parents and offspring belong to different classes. The covariance between parents and offspring is an **interclass** (between-class) covariance, while the covariance between collateral relatives is an **intraclass** (within-class) covariance. The analysis of variance (ANOVA), first proposed in Fisher’s 1918 paper, is used to estimate intraclass covariances.

Under the simplest ANOVA framework, we can consider the total variance of a trait to consist of two components: a **between-group** (also called the **among-group**) component (for example, differences in the mean values of different families) and a **within-group** component (the variation in trait value within each family). The total variance is the sum of the between and within group variances,

$$Var(T) = Var(B) + Var(W) \quad (3.1)$$

A key feature of ANOVA is that *the between-group variance equals the within-group covariance*. Thus, the larger the covariance between members of a family, the larger the fraction of total variation that is attributed to differences between family means.

To see this point, consider the following extreme patterns of phenotypes in full sib families:



### *Situation 1*

Suppose the between group variance  $Var(B) = 2.5$ , and the within-group variance  $Var(W) = 0.2$ . This gives a total phenotypic variance of  $V_P = Var(T) = Var(B) + Var(W) = 2.7$ . Here:

- members of a family resemble each other more closely than they do members of other families
- there are large differences in average phenotype between families

The resulting intraclass correlation  $t$  is

$$t = \frac{Cov(\text{full sibs})}{V_P} = \frac{Var(B)}{V_P} = 0.93$$

where we have used the above-mentioned ANOVA identity that the between-group variance equals the within-group covariance (here, the covariance between full sibs). Since elements of the same class are full-sibs, this is often denoted by  $t_{FS}$  to distinguish it from other intraclass correlations.

### Situation 2

Suppose the total (phenotypic) variance is the same as in situation 1, with  $Var(T) = V_P = 2.7$ . However, suppose there is no between-group variance ( $Var(B) = 0$ ), implying that  $Var(W) = 2.7$  and the intraclass correlation is  $t = 0$ . Here:

- members of a family resemble each other no more than they do members of other families
- there are no significant differences in average phenotype between families (i.e., all families have the same mean value)
- phenotypic resemblance is low, so genetic variation is low

Note that phenotypic resemblance among relatives can equivalently be considered as a measure of the *similarity* among a group of relatives for the phenotype of a quantitative trait (the covariance of family members), or the *difference* in phenotype between different families (the between-group variance = variance of family means).

## Causes of Phenotypic Covariance Among Relatives

Relatives resemble each other for quantitative traits more than they do unrelated members of the population for two potential reasons:

- relatives share genes. The closer the relationship, the higher the proportion of shared genes
- relatives share the same environment

### The Genetic Covariance Between Relatives

The Genetic Covariance  $Cov(G_x, G_y)$  = covariance of the genotypic values ( $G_x, G_y$ ) of the related individuals  $x$  and  $y$ .

We will first show how the genetic covariances between parent and offspring, full sibs, and half sibs depend on the genetic variances  $V_A$  and  $V_D$ . We will then discuss how these covariances are estimated in practice.

Genetic covariances arise because two related individuals are more likely to share alleles than are two unrelated individuals. Sharing alleles means having alleles that are **identical by descent** (IBD): namely that both copies of an allele can be traced back to a single copy in a recent common ancestor. Alleles can also be **identical in state** but not identical by descent. For example, both alleles in an  $A_1A_1$  individual are the same type (identical in state), but they are only identical by descent if both copies trace back to (descend from) a single copy in a recent ancestor.

For example, consider the offspring of two parents and label the four allelic copies in the parents by 1 - 4, independent of whether or not any are identical in state.

Parents:  $A_1A_2 \times A_3A_4$

Offspring:  $o_1 = A_1A_3$   $o_2 = A_1A_4$   $o_3 = A_2A_3$   $o_4 = A_2A_4$

Here,  $o_1$  and  $o_2$  share one allele IBD,  $o_1$  and  $o_3$  share two alleles IBD,  $o_1$  and  $o_4$  share no alleles IBD.

### 1. Offspring and one parent.

What is the covariance of genotypic values between an offspring ( $G_o$ ) and its parent ( $G_p$ )? Denoting the two parental alleles at a given locus by  $A_1A_2$ , since a parent and its offspring share *exactly* one allele, one allele in the offspring came from the parent (say  $A_1$ ), while the other offspring allele (denoted  $A_3$ ) came from the other parent. To consider the genetic contributions from a parent to its offspring, write the genotypic value of the parent as  $G_p = A + D$ . We can further decompose this by considering the contribution from each parental allele to the overall breeding value, with  $A = \alpha_1 + \alpha_2$ , and we can write the genotypic value of the parent as  $G_p = \alpha_1 + \alpha_2 + \delta_{12}$  where  $\delta_{12}$  denotes the dominance deviation for an  $A_1A_2$  genotype. Likewise, the genotypic value of its offspring is  $G_o = \alpha_1 + \alpha_3 + \delta_{13}$ , giving

$$Cov(G_o, G_p) = Cov(\alpha_1 + \alpha_2 + \delta_{12}, \alpha_1 + \alpha_3 + \delta_{13})$$

We can use the rules of covariances to expand this into nine covariance terms,

$$\begin{aligned} Cov(G_o, G_p) = & Cov(\alpha_1, \alpha_1) + Cov(\alpha_1, \alpha_3) + Cov(\alpha_1, \delta_{13}) \\ & + Cov(\alpha_2, \alpha_1) + Cov(\alpha_2, \alpha_3) + Cov(\alpha_2, \delta_{13}) \\ & + Cov(\delta_{12}, \alpha_1) + Cov(\delta_{12}, \alpha_3) + Cov(\delta_{12}, \delta_{13}) \end{aligned}$$

By the way have (intentionally) constructed  $\alpha$  and  $\delta$ , they are uncorrelated. Further,

$$Cov(\alpha_x, \alpha_y) = \begin{cases} 0 & \text{if } x \neq y, \text{ i.e., not IBD} \\ Var(A)/2 & \text{if } x = y, \text{ i.e., IBD} \end{cases} \quad (3.2a)$$

The last identity follows since  $Var(A) = Var(\alpha_1 + \alpha_2) = 2Var(\alpha_1)$ , so that

$$Var(\alpha_1) = Cov(\alpha_1, \alpha_1) = Var(A)/2$$

Hence, when individuals share one allele IBD, they share half the additive genetic variance. Likewise,

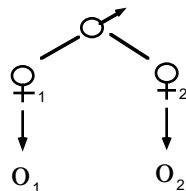
$$Cov(\delta_{xy}, \delta_{wz}) = \begin{cases} 0 & \text{if } xy \neq wz, \text{ i.e., both alleles are not IBD} \\ Var(D) & \text{if } xy = wz, \text{ both alleles are IBD} \end{cases} \quad (3.2b)$$

Two individuals only share the dominance variance when they share both alleles. Using the above identities (3.2a and b), eight of the above nine covariances are zero, leaving

$$Cov(G_o, G_p) = Cov(\alpha_1, \alpha_1) = Var(A)/2 \quad (3.3)$$

### 2. Half-sibs.

Here, one parent is shared, the other is drawn at random from the population;



The genetic covariance between half-sibs is the covariance of the genetic values between  $o_1$  and  $o_2$ .

To compute this, consider a single locus. First note that  $o_1$  and  $o_2$  share either one allele IBD (from the father) or no alleles IBD (since the mothers are assumed unrelated, these sibs cannot share both alleles IBD as they share no maternal alleles IBD). The probability that  $o_1$  and  $o_2$  both receive the same allele from the male is one-half (because whichever allele the male passes to  $o_1$ , the probability that he passes the same allele to  $o_2$  is one-half). In this case, the two offspring have one allele IBD, and the contribution to the genetic covariance when this occurs is  $Cov(\alpha_1, \alpha_1) = Var(A)/2$ . When  $o_1$  and  $o_2$  share no alleles IBD, they have no genetic covariance.

Summarizing:

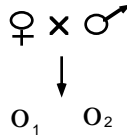
Case	Probability	Contribution
$o_1$ and $o_2$ have 0 alleles IBD	1/2	0
$o_1$ and $o_2$ have 1 allele IBD	1/2	$Var(A)/2$

giving the genetic covariance between half sibs as

$$Cov(G_{o_1}, G_{o_2}) = Var(A)/4 \quad (3.4)$$

### 3. Full-Sibs.

Both parents are in common,



What is the covariance of genotypic values of two full sibs?

Three cases are possible when considering pairs of full sibs: they can share either 0, 1, or 2 alleles IBD. Applying the same approach as for half sibs, if we can compute: 1) the probability of each case; and 2) the contribution to the genetic covariance for each case.

Each full sib receives one paternal and one maternal allele. The probability that each sib receives the same paternal allele is 1/2, which is also the probability each sib receives the same maternal allele. Hence,

$$\Pr(2 \text{ alleles IBD}) = \Pr(\text{paternal allele IBD}) \Pr(\text{maternal allele IBD}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$$\Pr(0 \text{ alleles IBD}) = \Pr(\text{paternal allele not IBD}) \Pr(\text{maternal allele not IBD}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$$\Pr(1 \text{ allele IBD}) = 1 - \Pr(2 \text{ alleles IBD}) - \Pr(0 \text{ alleles IBD}) = \frac{1}{2}$$

We saw above that when two relatives share one allele IBD, the contribution to the genetic covariance is  $Var(A)/2$ . When two relatives share both alleles IBD, each has the same genotype at the locus being considered, and the contribution is

$$Cov(\alpha_1 + \alpha_2 + \delta_{12}, \alpha_1 + \alpha_2 + \delta_{12}) = Var(\alpha_1 + \alpha_2 + \delta_{12}) = Var(A) + Var(D)$$

Putting these results together gives

Case	Probability	Contribution
$o_1$ and $o_2$ have 0 alleles IBD	1/4	0
$o_1$ and $o_2$ have 1 allele IBD	1/2	$Var(A)/2$
$o_1$ and $o_2$ have 2 allele IBD	1/4	$Var(A) + Var(D)$

This results in a genetic covariance between full sibs of

$$Cov(G_{o_1}, G_{o_2}) = \frac{1}{2} \frac{Var(A)}{2} + \frac{1}{4} (Var(A) + Var(D)) = \frac{Var(A)}{2} + \frac{Var(D)}{4} \quad (3.5)$$

#### 4. General degree of relationship.

Equations 3.2a and 3.2b suggest a general expression for the covariance between (noninbred) relatives, based on the probabilities that they share one and both alleles IBD.

If  $r_{xy} = (1/2) \text{Prob}(\text{relatives } x \text{ and } y \text{ have one allele IBD}) + \text{Prob}(\text{relatives } x \text{ and } y \text{ have both alleles IBD})$ , and  $u_{xy} = \text{Prob}(\text{relatives } x \text{ and } y \text{ have both alleles IBD})$ , then the genetic covariance between  $x$  and  $y$  is given by

$$Cov(G_x, G_y) = r_{xy}V_A + u_{xy}V_D \quad (3.6a)$$

If epistatic genetic variance is present, this can be generalized to

$$Cov(G_x, G_y) = r_{xy}V_A + u_{xy}V_D + r_{xy}^2V_{AA} + r_{xy}u_{xy}V_{AD} + u_{xy}^2V_{DD} + \dots \quad (3.6b)$$

#### Environmental Causes of Relationship Between Relatives

Shared environmental effects (such as a common maternal environment) also contribute to the covariance between relatives, and care must be taken to distinguish shared environmental effects from shared genetic effects.

If members of a family are reared together they share a common environmental value,  $E_c$ . If the common environmental circumstances are different for each family, the variance due to common environmental effects,  $V_{E_c}$ , causes greater similarity among members of a family, and greater differences among families, than would be expected just from the proportion of genes they share. Thus,  $V_{E_c}$  inflates the phenotypic covariance of sibs over what is expected from their genotypic covariance.

Just as we decomposed the total genotypic value into components, some shared, others not transmitted between relatives, we can do the same for environmental effects. In particular, we can write the total environmental effect  $E$  as the sum of a common environmental effect shared by the relatives  $E_c$ , a general environmental effect  $E_g$ , and a specific environmental effect  $E_s$  (unique to each individual). Hence, we can write  $E = E_c + E_g + E_s$ , partitioning the environmental variance as

$$V_E = V_{E_c} + V_{E_g} + V_{E_s} \quad (3.7)$$

We can further consider different possible sources of the common environmental effect  $E_c$ :

- $E_{cS}$  or  $E_{cL}$ : Shared effects due to sharing the space/location (different farms, cages)
- $E_{cT}$ : Temporal (changes in climactic or nutritional conditions over time)
- $E_{cM}$  Maternal (pre- and post-natal nutrition)

Thus, we can partition the environmental variance as

$$\begin{aligned} V_E &= V_{E_c} + V_{E_g} + V_{E_s} \\ &= V_{E_{cS}} + V_{E_{cT}} + V_{E_{cM}} + V_{E_c} + V_{E_g} + V_{E_s} \end{aligned}$$

Common environment effects mainly contribute to resemblance of sibs, but maternal environment effects can contribute to resemblance between mother and offspring as well.

$V_{E_{cS}}$  and  $V_{E_{cT}}$  can be eliminated, or estimated, by using the correct experimental design, but it is very difficult (except by cross-fostering) to eliminate or estimate  $V_{E_{cM}}$  from the covariance of full sibs. Further, cross-fostering only removes post-natal (past birth) maternal effects, it does not remove shared pre-natal maternal effects.

## Phenotypic Covariance Among Relatives and $h^2$

Summarizing the above results, the resulting covariances between common sets of relatives, the associated regression slopes ( $b$ ) or intra-class correlations ( $t$ ), and how these relate to estimates of  $h^2$  are as follows:

Relative Pair	$Cov$	$t$ or $b$	$h^2$
Parent-offspring ( $P-O$ )	$V_A/2$	$b_{O P} = \frac{1}{2} V_A/V_P$	$2b_{O P}$
Midparent-offspring ( $MP-O$ )	$V_A/2$	$*b_{O MP} = V_A/V_P$	$b_{O MP}$
Half-sibs (HS)	$V_A/4$	$t_{HS} = (1/4)V_A/V_P$	$4t_{HS}$
Full-sibs (FS)	$V_A/2 + V_D/4 + V_{Ec}$	$t_{FS} = \frac{V_A/2 + V_D/4 + V_{Ec}}{V_P}$	$2t_{FS} > h^2$

\*The midparent-offspring slope is computed as follows: using the properties of covariances,

$$\begin{aligned} Cov(O, MP) &= Cov(0, [P_f + P_m]/2) = \frac{Cov(0, P_f)}{2} + \frac{Cov(0, P_m)}{2} \\ &= \frac{V_A/2}{2} + \frac{V_A/2}{2} = V_A/2 \end{aligned}$$

The variance of the midparent values also follows from the properties of covariances, with

$$Var(MP) = Var\left(\frac{P_f + P_m}{2}\right) = \frac{Var(P_f)}{4} + \frac{Var(P_m)}{4} = V_P/2$$

The last equality assumes equal trait variances in both parents and that parental values are uncorrelated (i.e., no assortative mating). The regression slope equals the covariance between midparent and offspring divided by the midparent variance,

$$b_{O|MP} = \frac{Cov(O, MP)}{Var(MP)} = \frac{V_A/2}{V_P/2} = \frac{V_A}{V_P}$$



## Resemblance Between Relatives Problems

1. The *sm* (small) allele is a bristle mutation that segregates in an Australian *Drosophila* population, where the genotypic values for the wildtype (+) and small (*sm*) alleles are as follows: ++ : +*sm* : *sm sm* have values of 44 : 40 : 22. Suppose the environmental variance of bristle number is 6, and there are no common environmental effects due to maternal environment or rearing families together in vials. Assuming the *sm* locus is the only source of genetic variance, compute the regressions or intraclass correlations of bristle number between the following relatives:

- a: Offspring and midparent
- b: Half sibs
- c: Full sibs

Do the calculations for (i) populations where  $\text{freq}(sm) = 0.1$  and (ii) populations where  $\text{freq}(sm) = 0.9$ .

## Solutions to Resemblance Between Relatives Problems

1. Rescaling the genotypic values to  $a : d : -a$  gives  $11 : 7 : -11$ , or  $a = 11, d = 7$ . Hence, the genetic variances are given by

$$V_A = 2pq(a + d(q - p))^2 = 2pq(11 + 7(q - p))^2$$

and

$$V_D = (2pqd)^2 = q^2p^2 196$$

where  $q = \text{freq}(sm)$ . Hence:

$q$	$V_A$	$V_D$	$V_G$
0.1	5.25	1.59	6.84
0.9	49.60	1.59	51.19

Here  $V_E = 6$ , giving  $V_P = V_G + 6$ .

- a) Parent-offspring regression:  $b = V_A/V_P$ , or 0.41 for  $q = 0.1$ , 0.87 for  $q = 0.9$ .
- b) Half-sib correlation:  $t_{HS} = (1/4)V_A/V_P$ , or 0.10 for  $q = 0.1$ , 0.22 for  $q = 0.9$ .
- c) Full-sib correlation:  $t_{FS} = (V_A/2 + V_D/4)/V_P$ , or 0.24 for  $q = 0.1$ , 0.44 for  $q = 0.9$ .

## Solutions to Resemblance Between Relatives Problems

1. Suppose  $n$  offspring are measured to compute  $\bar{O}$ , the offspring mean for each parental pair.

a: What is  $Var(\bar{O})$ ?

$$\begin{aligned} Var(\bar{O}) &= Var\left(\frac{1}{n} \sum_{i=1}^n O_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^n O_i\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n Var(O_i) = \frac{n}{n^2} V_P = \frac{V_P}{n} \end{aligned}$$

b: What is the offspring mean - midparent covariance  $Cov(\bar{O}, MP)$ ?

$$\begin{aligned} Cov(\bar{O}, MP) &= Cov\left(\frac{1}{n} \sum_{i=1}^n O_i, \frac{P_f + P_m}{2}\right) = \frac{1}{2n} Cov\left(\sum_{i=1}^n O_i, P_f + P_m\right) \\ &= \frac{1}{2n} \sum_{i=1}^n Cov(O_i, P_f + P_m) = \frac{1}{2n} \sum_{i=1}^n Cov(O_i, P_f) + \sum_{i=1}^n Cov(O_i, P_m) \\ &= \frac{1}{2n} 2n Cov(O, P) = Cov(O, P) = V_A/2 \end{aligned}$$

c: What is the offspring mean - single parent covariance  $Cov(\bar{O}, P)$ ?

$$Cov(\bar{O}, P) = \frac{1}{n} \sum_{i=1}^n Cov(O_i, P) = \frac{1}{n} n Cov(O, P) = V_A/2$$

d: What is the slope of the offspring mean - midparent regression?

$$b_{\overline{O}|MP} = \frac{Cov(\bar{O}, MP)}{Var(MP)} + \frac{V_A/2}{V_P/2} = h^2$$

e: What is the slope of the offspring mean - single parent regression?

$$b_{\bar{O}|P} = \frac{Cov(\bar{O}, P)}{Var(P)} + \frac{V_A/2}{V_P} = \frac{h^2}{2}$$

f: What is the correlation between offspring mean and midparental value?

$$r(\bar{O}, MP) = \frac{Cov(\bar{O}, MP)}{\sqrt{Var(\bar{O}) \cdot Var(MP)}} = \frac{V_A/2}{\sqrt{V_P/n \cdot V_P/2}} = h^2 \sqrt{\frac{1}{2n}}$$