Quantitative characters II: heritability

The variance of a trait (x) is the average squared deviation of x from its mean:

$$V_P = \frac{1}{n} \sum (x - m_x)^2$$

This total phenotypic variance can be partitioned into components:

$$V_P = V_G + V_E$$ (genetic and environmental)

$$V_G = V_A + V_D + V_I$$ (additive, dom., interaction)

The broad-sense heritability is the fraction that’s genetic:

$$H^2 = \frac{V_G}{V_P}$$

The narrow-sense heritability is the fraction that’s additive genetic:

$$h^2 = \frac{V_A}{V_P}$$

$$h^2$$ determines (1) the resemblance of offspring to their parents, and (2) the population’s evolutionary response to selection.
$h^2$ is the regression (slope) of offspring on parents

$h^2 \approx 0$

$h^2 \approx \frac{1}{2}$

$h^2 \approx 1$

Definition of the regression coefficient (slope): $b_{yx} = \text{cov}(x,y)/\text{var}(x)$

Here $x$ is the midparent value (parental mean), $y$ is the offspring value (see Gillespie, Table 6.2).

The higher the slope, the better offspring resemble their parents.

In other words, the higher the heritability, the better offspring trait values are predicted by parental trait values.
This is why it's called "regression": offspring "regress" toward the mean!

Figure 6.5: The response to selection.
The “geometric” interpretation of heritability shows why $R = h^2 S$

$(h^2 = R / S)$

As it turns out, the additive genetic variance ($V_A$) is the part that makes offspring resemble their parents.
What’s the heritability of height in humans?

Scott Freeman and Jon Herron asked the students in their evolution course at the University of Washington to measure themselves and their parents.

Their regression plot is shown at the right.

The estimated heritability is 0.84.

That means 84% of the variance in height ($V_p$) is additive genetic variance ($V_A$).
The heritability of *genotypes* is 1.0 (illustrated by 30 Utah HapMap families).

One locus from each of eight x-somes. Genotypes coded 0/1/2.

<table>
<thead>
<tr>
<th>fam</th>
<th>chr1</th>
<th>chr2</th>
<th>chr7</th>
<th>chr8</th>
<th>chr19</th>
<th>chr20</th>
<th>chr21</th>
<th>chr22</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>21</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>22</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>26</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>27</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>29</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>7</td>
</tr>
</tbody>
</table>

Each row is a family. The sum of genotype scores ("plus" alleles) is shown in the last set of columns.
The average regression is close to 1, as is that of the sum over loci.

<table>
<thead>
<tr>
<th>Chr 1 (A/C at 53433581)</th>
<th>Chr 2 (G/T at 224346716)</th>
<th>Chr 7 (A/T at 92947019)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0: 4 5 4</td>
<td>2.0: 2 1</td>
<td>2.0: 8 10</td>
</tr>
<tr>
<td>1.0: 2 6 8</td>
<td>1.0: 6 4</td>
<td>1.0: 2 3 7</td>
</tr>
<tr>
<td>0.0: 1</td>
<td>0.0: 9 7 1</td>
<td>0.0: 0 0.5 1 1.5 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>r = 0.672  b = 0.792</td>
</tr>
<tr>
<td>0 0.5 1 1.5 2</td>
<td>0 0.5 1 1.5 2</td>
<td></td>
</tr>
<tr>
<td>r = 0.409  b = 0.569</td>
<td>r = 0.731  b = 1.200</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chr 8 (A/G at 122870354)</th>
<th>Chr 19 (A/G at 48986689)</th>
<th>Chr 20 (A/G at 48392908)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0: 2 7 8</td>
<td>2.0: 2 3 1</td>
<td>2.0: 3 1</td>
</tr>
<tr>
<td>1.0: 5 7</td>
<td>1.0: 11 3</td>
<td>1.0: 7 3 2</td>
</tr>
<tr>
<td>0.0: 1</td>
<td>0.0: 5 5</td>
<td>0.0: 7 6 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0.5 1 1.5 2</td>
<td>0 0.5 1 1.5 2</td>
<td>0 0.5 1 1.5 2</td>
</tr>
<tr>
<td>r = 0.569  b = 0.875</td>
<td>r = 0.822  b = 1.216</td>
<td>r = 0.682  b = 1.048</td>
</tr>
</tbody>
</table>

The regressions (b) are greater than the correlations (r) because \( \text{var}(x) \) (i.e., of midparent values) < \( \text{var}(y) \) (i.e., of offspring values). See Gillespie, table 6.2.
Chr 21 (C/T at 18426726)

2.0:  1  5  4  
1.0:  3  7  7  
0.0:  1  2  
0  0.5  1  1.5  2  
r = 0.629  b = 0.875

Chr 22 (A/G at 16361045)

2.0:  1  5  9  
1.0:  6  8  
0.0:  1  
0  0.5  1  1.5  2  
r = 0.673  b = 1.006

Sum of genotypic values for all 8 loci

11.0:  1  1  2  4  
10.0:  2  3  2  1  1  
9.0:  1  2  1  
8.0:  1  1  1  1  
7.0:  1  1  1  
6.0:  1  1  
6  7  8  9  10  11  
r = 0.651  b = 1.007

Expected: $b = h^2 = 1$
$r = h^2 / \sqrt{2} = 1 / 1.414 = 0.71$
What about the variation induced by environmental factors? After all, even clones and identical twins differ from each other!

Clones (cuttings) of *Achillea* grown at three different elevations where the species normally occurs in California.

Edward East’s *Nicotiana* plants growing in the same garden plots.

Leaves from a natural clone of quaking aspen (*Populus tremuloides*) growing at the top of Millcreek Canyon.
Total phenotypic variance = genetic variance + environmental variance

What you see is what you get from two distinct sources that can be separated.

1. **Genetic variance** is the variance among *phenotypes* caused by genotypic differences among individuals (holding their environments constant).

2. **Environmental variance** is the variance among *phenotypes* caused by differences in the experiences of individuals (holding genotypes constant).

**Example:** Suppose the average trait values of AA, Aa and aa individuals are -1, 0, and +1 units, and \( p = q = 0.5 \).

Then the *genetic variance* (average squared deviation from the population mean) is 0.5.

But suppose 25% of each genotype deviates one unit above or below its average trait value, because of the environment.

Then the *environmental variance* is also 0.5.

The resulting *phenotypic variance* is \( 0.5 + 0.5 = 1.0 \).

\[
V_P = V_G + V_E = (V_A + V_D + V_I) + (V_E)
\]

The trait’s *heritability* is the fraction of \( V_P \) that is *genetic* (actually, *additive genetic*, as we will see).
Not all genetic variance is additive! (A silly but instructive model.)

Consider a simple quantitative trait \( x \) controlled in a symmetrically overdominant manner by two alleles at one locus.

Assume that there is no environmental variance.

<table>
<thead>
<tr>
<th>genotype</th>
<th>phenotype ((x))</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>0</td>
<td>( p^2 )</td>
</tr>
<tr>
<td>Aa</td>
<td>1</td>
<td>( 2p(1-p) )</td>
</tr>
<tr>
<td>aa</td>
<td>0</td>
<td>( (1-p)^2 )</td>
</tr>
</tbody>
</table>

Thus \( \bar{x} = 2p(1-p) \), and \( \text{Var}(x) = V_P = V_G = E(x^2) - [E(x)]^2 = 2p(1-p)[1-2p(1-p)] \).

What happens if we select for higher values of \( x \)?
The heritability disappears when the genetic variance is greatest!

Now we can calculate the heritability, which is

\[ h^2 = \frac{V_A}{V_p} = \frac{(1-2p)^2}{[1-2p(1-p)]} \]

\[ p \approx 0 \text{ or } 1 \]

Not much variance, but \( h^2 \approx 1 \)

\[ p = 0.5 \]

Lots of variance, but \( h^2 = 0! \)

At \( p = 0.5 \), all of the genetic variance is \textit{dominance variance}, not additive variance.
Dominance variance arises from non-additive relationships between the “dosage” of an allele (number carried) and the resulting phenotype.

No dominance. Phenotypes: $A_1A_1 = 1$; $A_1A_2 = 1.5$; $A_2A_2 = 2$

Complete dominance. Phenotypes: $A_1A_1 = 1$; $A_1A_2 = 2$; $A_2A_2 = 2$
How to estimate the components of the phenotypic variance ($V_P$)

1. Measure phenotypes (trait values) in a large random sample of the population.
2. Calculate the mean and variance: the variance is $V_P$.
3. Estimate the heritability, either of two ways:
   (a) regress offspring on midparent values
   (b) measure the response to selection:
   \[ h^2 = R/S \]
4. The additive variance ($V_A$) is the heritable fraction of the total: $V_A = h^2 V_P$.
5. The remainder is environmental ($V_E$) and dominance variance (and other minor stuff).
6. If we can clone or closely inbreed members of the species, or find identical twins, then we can directly estimate the environmental variance.
Leaf shape within and among six quaking aspen clones

<table>
<thead>
<tr>
<th>Location</th>
<th>Clone 1</th>
<th>Clone 2</th>
<th>Clone 3</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Canyon</td>
<td>0.902</td>
<td>0.992</td>
<td>1.075</td>
<td>0.963</td>
<td>0.00990</td>
</tr>
<tr>
<td>Clone 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clone 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clone 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper Millcreek</td>
<td>0.861</td>
<td>1.028</td>
<td>0.918</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clone 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clone 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clone 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Leaf length / leaf width

0.7 0.8 0.9 1.0 1.1 1.2

mean variance
Analysis of variance (ANOVA)

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clone 1</td>
<td>0.902</td>
<td>0.00351</td>
</tr>
<tr>
<td>Clone 2</td>
<td>0.992</td>
<td>0.00237</td>
</tr>
<tr>
<td>Clone 3</td>
<td>1.075</td>
<td>0.00271</td>
</tr>
<tr>
<td>Clone 1</td>
<td>0.861</td>
<td>0.00552</td>
</tr>
<tr>
<td>Clone 2</td>
<td>1.028</td>
<td>0.00200</td>
</tr>
<tr>
<td>Clone 3</td>
<td>0.918</td>
<td>0.00947</td>
</tr>
<tr>
<td>All</td>
<td>0.963</td>
<td>0.00990</td>
</tr>
</tbody>
</table>

Variance among clones
= \text{var}(0.902, 0.992, \ldots, 0.918)
= 0.00564

Variance within clones
= \text{mean}(0.00351, \ldots, 0.00947)
= 0.00426

Total variance
= 0.00564 + 0.00426
= 0.00990

Fraction explained by clones
= 0.00564 / 0.00990
= 0.57 = H^2
7. The dominance variance can be separated from the additive variance by exploiting the different ways these components appear in the covariances between different kinds of relatives. For example:

\[
\begin{align*}
\text{cov(parent-offspring)} & \approx \frac{1}{2} V_A \\
\text{cov(half sibs)} & \approx \frac{1}{4} V_A \\
\text{cov(full sibs)} & \approx \frac{1}{2} V_A + \frac{1}{4} V_D
\end{align*}
\]

See the Quantitative Traits lecture notes, and Gillespie, for more on this.

And it works in the “pure overdominance” model developed above...
No heritability, but a **positive correlation between sibs**, in the pure-overdominance model with \( p = 0.5 \).

<table>
<thead>
<tr>
<th>Parents</th>
<th>Offspring Phenotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>both 0</td>
<td></td>
</tr>
<tr>
<td>[0,0]</td>
<td></td>
</tr>
<tr>
<td><strong>parents</strong></td>
<td><strong>Freq</strong></td>
</tr>
<tr>
<td>AA x AA</td>
<td>0.0625</td>
</tr>
<tr>
<td>AA x Aa</td>
<td>0.25</td>
</tr>
<tr>
<td>AA x aa</td>
<td>0.125</td>
</tr>
<tr>
<td>Aa x Aa</td>
<td>0.25</td>
</tr>
<tr>
<td>Aa x aa</td>
<td>0.25</td>
</tr>
<tr>
<td>aa x aa</td>
<td>0.0625</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\text{var} &= 0.25, \quad \text{cov} = 0 \\
\frac{\text{cov}}{\text{var}} &= b = 0
\end{align*}
\]
7. The dominance variance can be separated from the additive variance by exploiting the different ways these components appear in the covariances between different kinds of relatives. For example:

\[
\text{cov(}\text{parent-offspring}) \approx \frac{1}{2} V_A
\]
\[
\text{cov(}\text{half sibs}) \approx \frac{1}{4} V_A
\]
\[
\text{cov(}\text{full sibs}) \approx \frac{1}{2} V_A + \frac{1}{4} V_D \quad \text{(see QT lecture notes)}
\]

An interesting and general finding is that traits closely related to fitness tend to have little additive variance but more dominance and interaction variance (epistasis) than typical morphological traits.

What might be the explanation?

<table>
<thead>
<tr>
<th>Character</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bristles</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thorax</td>
<td>52</td>
<td>43</td>
<td>30</td>
<td>18</td>
</tr>
<tr>
<td>Ovary</td>
<td>9</td>
<td>6</td>
<td>40</td>
<td>44</td>
</tr>
<tr>
<td>Eggs</td>
<td>39</td>
<td>51</td>
<td>30</td>
<td>38</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characters:</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Number of bristles on 4th + 5th abdominal segments (Clayton, Morris, and Robertson, 1957; Reeve and Robertson, 1954).</td>
</tr>
<tr>
<td>(2) Length of thorax (F. W. Robertson, 1957b).</td>
</tr>
<tr>
<td>(3) Size of ovaries, i.e. number of ovarioles in both ovaries. (F. W. Robertson, 1957a).</td>
</tr>
<tr>
<td>(4) Number of eggs laid in 4 days (4th to 8th after emergence) (F. W. Robertson, 1957b).</td>
</tr>
</tbody>
</table>
Summary

The narrow-sense heritability of a trait is the fraction of the total phenotypic variance that is caused by the additive effects of genes.

There can be considerable non-additive genetic variance, but this does not contribute to the resemblance between parents and offspring, or the response to selection.

(But the dominance variance increases the resemblance of full siblings.)

There can also be a lot of “environmental variance” (that is, variance of the trait values that is caused by effects of the environment).

These three components of the phenotypic variance literally add up to the total: \( V_p = V_A + (V_D + V_I) + V_E \)

The analysis of variance (ANOVA) was originally invented to allow these components to be estimated separately.