# STUDIES IN HUMAN INHERITANCE. X. 

# A TABLE TO DETERMINE THE PROPORTION OF RECESSIVES TO BE EXPECTED IN VARIOUS MATINGS INVOLVING A UNIT CHARACTER* 

LAURENCE H. SNYDER<br>Ohio State University, Columbus, Ohio

Received February 9, 1933
In genetic studies on animals and plants, the proof of the allelomorphic nature of two genes, one of which is dominant, has depended largely on the obtaining of the $3: 1$ ratio from the matings of known heterozygotes or of the $1: 1$ ratio from the backcross. Heterozygotes have been relatively easy to determine in such cases by using $\mathrm{F}_{1}$ individuals, or by making test matings. When we approach the study of human heredity, however, we are unable to make test matings, and we are faced with considerable difficulty in determining whether an individual exhibiting a dominant trait is homozygous or heterozygous. In some instances this may be determined from a knowledge of the individual's parents or offspring. In many cases, however, our data are limited to two generations. Furthermore, a human family is never large enough to prove that an individual showing a dominant trait and producing no recessive offspring is really homozygous.

It becomes necessary, therefore, in the study of human heredity, to use an analysis which obviates the need of distinguishing between homozygous and heterozygous dominants. In a random series of matings where both parents show a dominant character, the offspring will show dominance and recessiveness, respectively, in a proportion greater than $3: 1$, because of the fact that some of the dominant parents are homozygous, and will produce all dominant offspring. Similarly, in a random series of matings in which one parent shows the dominant character, the other the recessive, the ratio of dominance to recessiveness among the offspring will be greater than $1: 1$.

The exact proportions of recessives to be expected in these types of matings may, however, be predicted very exactly by an analysis of the genes on a frequency basis. These proportions will vary among various factors, depending on the relative frequencies of the allelomorphic genes in the population, and will vary for the same factor among different races, if the frequencies of the genes vary from race to race.

The expected proportions of recessive offspring from the matings of

[^0]dominants with dominants, and dominants with recessives, respectively, are calculated as follows:
Let $p=$ frequency of the dominant gene of a pair of allelomorphs, and let $q=$ frequency of the recessive gene.

Then $\mathrm{p}+\mathrm{q}=1$. It is readily seen that
$\mathrm{p}^{2}+2 \mathrm{pq}=$ the dominant individuals of the population. (A)
$\mathrm{q}^{2}=$ the recessive individuals of the population. (B)
$\mathrm{q}=\sqrt{\mathrm{B}}$
$\mathrm{p}=1-\sqrt{\mathrm{B}}$
A relationship may be readily demonstrated between the frequencies of the allelomorphic genes and the proportions of recessive offspring to be expected in random matings of dominants with dominants, and dominants with recessives.

Since $\mathrm{p}^{2}=$ homozygous dominants and $2 \mathrm{pq}=$ heterozygous dominants,

$$
\frac{\mathrm{p}^{2}}{\mathrm{p}^{2}+2 \mathrm{pq}}=\text { proportion of dominants which are homozygous }
$$

$$
\frac{2 \mathrm{pq}}{\mathrm{p}^{2}+2 \mathrm{pq}}=\text { proportion of dominants which are heterozygous }
$$

The only recessive offspring produced in matings of dominants with dominants will be one-quarter of the offspring of matings of heterozygous dominants with heterozygous dominants. Similarly, the only recessive offspring produced in matings of dominants with recessives will be one-half of the offspring of the matings of heterozygous dominants with recessives. Formulae for these proportions are derived as follows:
Let R = proportion of recessive offspring to be expected from matings of dominants with dominants,
And $\mathrm{S}=$ proportion of recessive offspring to be expected from matings of dominants with recessives.

Then

$$
\mathrm{R}=1 / 4\left(\frac{2 \mathrm{pq}}{\mathrm{p}^{2}+2 \mathrm{pq}}\right)^{2}=\left(\frac{\mathrm{q}}{\mathrm{p}+2 \mathrm{q}}\right)^{2}
$$

and

$$
S=1 / 2\left(\frac{2 p q}{p^{2}+2 p q}\right)=\frac{q}{p+2 q}
$$

Since the results of these formulae will vary depending on the frequencies of the genes concerned, which in turn are derived from the proportion of recessive individuals in the general population, the following table has been prepared. This table gives directly from the observed proportion of recessive individuals in the population, the expected proportion of recessive offspring from matings of dominants with dominants, and dominants
with recessives. The calculations were done to six decimal places with an electric Marchant calculator, and the results are presented in the table to the nearest fourth decimal place.

While the table will be particularly useful in studies of human heredity, it will be found applicable to studies of animals and plants, where it is impractical or impossible to determine the heterozygosity of dominants. It is not necessary, for example, to make the general statement so often encountered in genetic studies, that "from certain matings of dominant with dominant, a ratio approaching $3: 1$ was obtained."

As an example of the use of the table, it is found that among white Americans, .298 are unable to taste phenyl-thio-carbamide (SNyder 1932, Ohio Journal of Science 32:436). This taste deficiency appears on simple inspection of the family histories, to be an autosomal recessive character. To further prove the unit character nature of the deficiency, the table indicates that in matings of tasters with tasters (Column R) we should expect .1247 of the offspring to be recessive, that is, taste-deficient. From a study of 800 families we find that in such matings .1228 of the offspring are recessive, a difference of $.0019 \pm .007$. Similarly, in matings of tasters with non-tasters, we find (Column S), that we should expect .3531 of the offspring to be recessive if we are really dealing with a unit character. The proportion actually found is .3653 , a difference of $.0122 \pm .012$.

The probable errors of the observed and calculated proportions are as follows:

For the P.E. of the observed proportions, the well-known formula $.6745 \sqrt{\frac{\mathrm{xy}}{\mathrm{n}}}$ is used, where $\mathrm{x}=$ observed proportion of dominant children, $\mathrm{y}=$ observed proportion of recessive children, and $\mathrm{n}=$ total number of offspring, for each type of mating.

For the calculated proportions, where $\mathrm{N}=100$ or more, the formulae are

$$
\begin{aligned}
& \text { P.E. }\left(\frac{q}{p+2 q}\right)^{2}=\frac{.6745(1-\sqrt{b})}{(1+\sqrt{b})^{2}} \sqrt{\frac{b}{N(1-b)}} \\
& \text { P.E. } \frac{q}{p+2 q}=\frac{.6745(1-\sqrt{b})}{2(1+\sqrt{b})} \sqrt{\frac{1}{N(1-b)}}
\end{aligned}
$$

where $N=$ Total number of individuals tested in deriving $A$ and $B$. In practice, $b$, which is the true value of the proportion of recessives in the population, must be taken as $B$, the observed value from the complete sample of individuals studied, just as in the foregoing familiar formula $.6745 \sqrt{\frac{x y}{n}}, x$ and $y$ actually represent the true proportions, but are in practice taken as the observed proportions.

Since $p$ and $q$ are subject to error because they are derived from A and B, which are themselves representative only of a sample of the whole population, the derivation of these formulae is as follows:
derivation of a formula for the probable error of $\frac{q}{p+2 q}$
Let $b=$ true value of the proportion of recessives in the population.
And $B=$ the value of $b$ obtained from a sample of the population.
Then $B=b+e$
In which $\mathrm{e}=$ the difference between the value of b calculated from the sample and the true value of $b$.
Since True $q=\sqrt{b}$
and True $p=1-\sqrt{b}$
Then $\operatorname{True} \frac{q}{p+2 q}=\frac{\sqrt{\mathbf{b}}}{1+\sqrt{\bar{b}}}$
But the value of $b$ calculated from a sample may be in error by an amount e so that the obtained value of $q=\sqrt{b+e}$
Hence calculated $\frac{q}{p+2 q}$ may be in error by an amount equal to Calcu-
lated $\frac{q}{p+2 q}-$ True $\frac{q}{p+2 q}$ or E, the errorin $\frac{q}{p+2 q}=\frac{\sqrt{b+e}}{1+\sqrt{b+e}}-\frac{\sqrt{b}}{1+\sqrt{b}}$ (1)
Hence,

$$
\begin{equation*}
E=\frac{1}{1+\sqrt{b}}-\frac{1}{1+\sqrt{\mathrm{b}+\mathrm{e}}} \tag{2}
\end{equation*}
$$

$$
\frac{1}{1+\sqrt{b+e}}=\frac{1}{1+\sqrt{b}}-E \text { or } \frac{1}{1+\sqrt{b+e}}=\frac{1-(1+\sqrt{b}) E}{1+\sqrt{b}}
$$

Inverting,

$$
1+\sqrt{b+e}=\frac{1+\sqrt{b}}{1-(1+\sqrt{b}) E} \text { or } \sqrt{b+e}=
$$

$$
\begin{gather*}
\frac{(1+\sqrt{\mathrm{b}})-1+(1+\sqrt{\mathrm{b}}) \mathrm{E}}{1-(1+\sqrt{\mathrm{b}}) \mathrm{E}} \\
\sqrt{\mathrm{~b}+\mathrm{e}}=\frac{\frac{\sqrt{\mathrm{b}}}{\frac{1+\sqrt{\mathrm{b}}}{}+\mathrm{E}}}{\frac{1}{1+\sqrt{\mathrm{b}}}-\mathrm{E}} \tag{3}
\end{gather*}
$$

Hence,

For simplicity in symbols we may now let $\frac{1}{1+\sqrt{\mathrm{b}}}=\mathrm{k}$
Equation (3) then becomes $\sqrt{b+e}=\frac{k \sqrt{b}+E}{k-E}$

Now, to eliminate E in the denominator, both numerator and denominator of the right side of equation (4) are multiplied by $\mathrm{k}^{3}+\mathrm{k}^{2} \mathrm{E}+\mathrm{kE}^{2}+\mathrm{E}^{3}$. The equation resulting is:

$$
\begin{equation*}
\sqrt{\mathrm{b}+\mathrm{e}}=\frac{\mathrm{k}^{4} \sqrt{\mathrm{~b}}+\mathrm{k}^{2} \mathrm{E}+\mathrm{kE}^{2}+\mathrm{E}^{3}+\mathrm{E}^{4}}{\mathrm{k}^{4}-\mathrm{E}^{4}} \tag{5}
\end{equation*}
$$

Squaring both sides of Equation (5)

$$
\begin{align*}
& b+e=\left(\frac{1}{k^{8}-2 k^{4} E^{4}+E^{8}}\right)\left(k^{8} b+k^{4} E^{2}+k^{2} E^{4}+E^{6}+E^{8}+2 k^{6} \sqrt{b} E\right. \\
& +2 k^{5} \sqrt{b} E^{2}+2 k^{4} \sqrt{b} E^{3}+2 k^{4} \sqrt{b} E^{4}+2 k^{3} E^{3}+2 k^{2} E^{4}+2 k^{2} E^{5} \\
&  \tag{6}\\
& \left.+2 k E^{5}+2 k E^{6}+2 E^{7}\right)
\end{align*}
$$

However, the value of E cannot exceed .5 and in case the sample from which b is computed includes 100 or more cases, it is highly improbable that $E$ will exceed . 03 . Therefore, powers of $E$ higher than $E^{3}$ may be disregarded since their values approximate zero. Equation (6) then becomes:

$$
\begin{equation*}
\mathrm{b}+\mathrm{e}=\frac{1}{\mathrm{k}^{8}}\left(\mathrm{k}^{8} \mathrm{~b}+2 \mathrm{k}^{6} \sqrt{\mathrm{~b}} \mathrm{E}+\mathrm{k}^{4} \mathrm{E}^{2}+2 \mathrm{k}^{5} \sqrt{\mathrm{~b}} \mathrm{E}^{2}+2 \mathrm{k}^{4} \sqrt{\mathrm{~b}} \mathrm{E}^{3}+2 \mathrm{k}^{3} \mathrm{E}^{3}\right) \tag{7}
\end{equation*}
$$

Hence, $\quad e=\frac{2 \sqrt{b}}{k^{2}} E+\left(\frac{1}{k^{4}}+\frac{2 \sqrt{\mathrm{~b}}}{\mathrm{k}^{3}}\right) \mathrm{E}^{2}+\left(\frac{2 \sqrt{\mathrm{~b}}}{\mathrm{k}^{4}}+\frac{2}{\mathrm{k}^{5}}\right) \mathrm{E}^{3}$.
Squaring, $\quad e^{2}=\frac{4 b}{k^{4}} \mathrm{E}^{2}+\left(\frac{1}{\mathrm{k}^{4}}+\frac{2 \sqrt{\mathrm{~b}}}{\mathrm{k}^{3}}\right)^{2} \mathrm{E}^{4}+\left(\frac{2 \sqrt{\mathrm{~b}}}{\mathrm{k}^{4}}+\frac{2}{\mathrm{k}^{5}}\right)^{2} \mathrm{E}^{6}$

$$
\begin{aligned}
& +\frac{4 \sqrt{b}}{k^{2}}\left(\frac{1}{k^{4}}+\frac{2 \sqrt{b}}{k^{3}}\right) \mathrm{E}^{3}+\frac{4 \sqrt{b}}{\mathrm{k}^{2}}\left(\frac{2 \sqrt{ } b}{k^{4}}+\frac{2}{k^{5}}\right) \mathrm{E}^{4} \\
& +2\left(\frac{1}{k^{4}}+\frac{2 \sqrt{b}}{k^{3}}\right)\left(\frac{2 \sqrt{b}}{k^{4}}+\frac{2}{k^{5}}\right) E^{5}
\end{aligned}
$$

Summing this expression and disregarding values of $\sum \mathrm{E}^{3}, \sum \mathrm{E}^{4}, \sum \mathrm{E}^{5}$, $\sum \mathrm{E}^{6}$, all of which approximate zero, we have,

$$
\sum \mathrm{e}^{2}=\frac{4 \mathrm{~b}}{\mathrm{k}^{4}} \sum \mathrm{E}^{2}
$$

But, since $e$ is the error in estimating the proportion of $b$ in the population, the standard deviation of $e=\sqrt{\frac{a b}{N}}$.

Then

$$
\sum \mathrm{e}^{2}=\mathrm{N}\left(\sqrt{\frac{\mathrm{ab}}{\mathrm{~N}}}\right)^{2}=\mathrm{ab}
$$

Hence, $\quad \sum E^{2}=\frac{k^{4} a b}{4 b}=\frac{k^{4} a}{4}=\frac{a}{4(1+\sqrt{\bar{b}})^{4}}=\frac{1-\sqrt{b}}{4(1+\sqrt{\bar{b}})^{3}}$.
Now, to evaluate $\sum \mathrm{E}$, let us go back to equation (7).
Summing this expression and disregarding values of $\sum \mathrm{E}^{3}$, which approximate zero, we have:

$$
\sum \mathrm{e}=\frac{2 \sqrt{\mathrm{~b}}}{\mathrm{k}^{2}} \sum \mathrm{E}+\left(\frac{1}{\mathrm{k}^{4}}+\frac{2 \sqrt{\mathrm{~b}}}{\mathrm{k}^{3}}\right) \sum \mathrm{E}^{2}
$$

But the sum of the errors in estimating $b$ may be assumed to be zero, hence

$$
\frac{2 \sqrt{\mathrm{~b}}}{\mathrm{k}^{2}} \sum \mathrm{E}=-\left(\frac{1}{\mathrm{k}^{4}}+\frac{2 \sqrt{\mathrm{~b}}}{\mathrm{k}^{3}}\right) \sum \mathrm{E}^{2}
$$

Therefore

$$
\sum \mathrm{E}=-\left(\frac{1}{2 \mathrm{k}^{2} \sqrt{\mathrm{~b}}}+\frac{1}{\mathrm{k}}\right) \sum \mathrm{E}^{2}
$$

Since

$$
\sum \mathrm{E}^{2}=\frac{\mathrm{k}^{4} \mathrm{a}}{4}
$$

then

$$
\begin{equation*}
\sum E=-\left(\frac{1}{2 k^{2} \sqrt{b}}+\frac{1}{k}\right) \frac{k^{4} a}{4}=-\frac{k^{2} a}{8 \sqrt{b}}-\frac{k^{3} a}{4} \tag{9}
\end{equation*}
$$

The standard deviation of $\mathrm{E}, \sigma \mathrm{E}=\sqrt{\frac{\Sigma \mathrm{E}^{2}}{\mathrm{~N}}-\left(\frac{\Sigma \mathrm{E}}{\mathrm{N}}\right)^{2}}$.
Hence,

$$
\begin{aligned}
\sigma \mathrm{E} & =\sqrt{\frac{\mathrm{k}^{4} \mathrm{a}}{4 \mathrm{~N}}-\left(-\frac{\mathrm{k}^{2} a}{8 \sqrt{\bar{b} N}}-\frac{\mathrm{k}^{3} a}{4 N}\right)^{2}} \\
& =\sqrt{\frac{16 \mathrm{k}^{4} a b N-\mathrm{k}^{4} a^{2}-4 k^{5} a^{2} \sqrt{b}-4 k^{6} a^{2} b}{64 b N^{2}}}
\end{aligned}
$$

Since $k=\frac{1}{1+\sqrt{b}}$, and $a=(1-\sqrt{b})(1+\sqrt{\bar{b}})$, then

$$
\sigma E=\frac{1}{8 N(1+\sqrt{b})}
$$

$$
\sqrt{\frac{16 N(1-\sqrt{b})}{1+\sqrt{b}}-\frac{(1-\sqrt{b})^{2}}{b}-\frac{4(1-\sqrt{b})^{2}}{(1+\sqrt{b})^{2} \sqrt{b}}-\frac{4(1-\sqrt{b})^{2}}{(1+\sqrt{b})^{2}}}
$$

Then the P.E. of $\frac{q}{p+2 q}=\frac{.6745}{8 N(1+\sqrt{b})}$

$$
\begin{equation*}
\sqrt{\frac{16 N(1-\sqrt{b})}{1+\sqrt{b}}-\frac{(1-\sqrt{b})^{2}}{b}-\frac{4(1-\sqrt{b})^{2}}{(1+\sqrt{b})^{2} \sqrt{b}}-\frac{4(1-\sqrt{b})^{2}}{(1+\sqrt{b})^{2}}} \tag{10}
\end{equation*}
$$

Since the sum of the last three terms under the radical is very small in
relation to the first term when N is 100 or more the last three terms may be omitted for approximate calculations.
In this case, P.E. of $\frac{q}{p+2 q}=\frac{.6745(1-\sqrt{b})}{2(1+\sqrt{b})} \sqrt{\frac{1}{(1-b) N}}$.
DERIVATION OF A FORMULA FOR THE PROBABLE ERROR OF $\left(\frac{q}{p+2 q}\right)^{2}$
In the previous section a formula for the probable error of $\frac{q}{p+2 q}$ was developed. Using the same symbols as before the calculated value of $\left(\frac{q}{p+2 q}\right)^{2}$ may be in error by an amount equal to: Calculated $\left(\frac{q}{p+2 q}\right)^{2}$ - True $\left(\frac{q}{p+2 q}\right)^{2}$.

Hence, $G$, the error in $\left(\frac{q}{p+2 q}\right)^{2}=\left(\frac{\sqrt{b+e}}{1+\sqrt{b+e}}\right)^{2}-\left(\frac{\sqrt{b}}{1+\sqrt{b}}\right)^{2}$.
But since

$$
\begin{equation*}
\mathrm{E}=\frac{\sqrt{\mathrm{b}+\mathrm{e}}}{1+\sqrt{\mathrm{b}+\mathrm{e}}}-\frac{\sqrt{\mathrm{b}}}{1+\sqrt{\mathrm{b}}} \tag{11}
\end{equation*}
$$

(Equation (1))
then

$$
\begin{equation*}
G=E^{2}+\frac{2 \sqrt{b}}{1+\sqrt{b}} E \tag{12}
\end{equation*}
$$

Summing, $\quad \sum G=\sum E^{2}+\frac{2 \sqrt{b}}{1+\sqrt{b}} \sum E$.
Since

$$
\begin{equation*}
\sum \mathrm{E}^{2}=\frac{\mathrm{k}^{4} \mathrm{a}}{4} \tag{8}
\end{equation*}
$$

and

$$
\begin{equation*}
\sum \mathrm{E}=-\frac{\mathrm{k}^{2} \mathrm{a}}{8 \sqrt{\mathrm{~b}}}-\frac{\mathrm{k}^{3} \mathrm{a}}{4} \tag{9}
\end{equation*}
$$

and

$$
\mathrm{k}=\frac{1}{1+\sqrt{\mathrm{b}}}
$$

then

$$
\begin{equation*}
\sum G=\frac{k^{4} a}{4}(1-2 \sqrt{b})-\frac{k^{3} a}{4} \tag{13}
\end{equation*}
$$

Squaring Equation (12): $\quad G^{2}=E^{4}+\frac{4 \sqrt{b}}{1+\sqrt{b}} E^{3}+\frac{4 b}{(1+\sqrt{b})^{2}} E^{2}$.
Summing, $\quad \sum G^{2}=\sum E^{4}+\frac{4 \sqrt{\mathrm{~b}}}{1+\sqrt{\mathrm{b}}} \sum \mathrm{E}^{3}+\frac{4 \mathrm{~b}}{(1+\sqrt{\mathrm{b}})^{2}} \sum \mathrm{E}^{2}$.

However, we may disregard values of $\sum \mathrm{E}^{3}$ and $\sum \mathrm{E}^{4}$ since they approximate zero.

Hence,

$$
\begin{equation*}
\sum \mathrm{G}^{2}=\frac{4 \mathrm{~b}}{(1+\sqrt{\mathrm{b}})^{2}} \sum \mathrm{E}^{2}=\mathrm{k}^{6} \mathrm{ab} \tag{15}
\end{equation*}
$$

The standard deviation of $\mathrm{G}, \sigma \mathrm{G}=\sqrt{\frac{\sum \mathrm{G}^{2}}{\mathrm{~N}}-\left(\frac{\sum \mathrm{G}}{\mathrm{N}}\right)^{2}}$.
Therefore, $\quad \sigma_{G}=\sqrt{\frac{k^{6} a b}{N}-\left(\frac{\frac{k^{4} a}{4}(1-2 \sqrt{b})-\frac{k^{3} a}{4}}{N}\right)^{2}}$

$$
\begin{equation*}
\sigma_{G}=\frac{1-\sqrt{\mathrm{b}}}{4 \mathrm{~N}(1+\sqrt{\mathrm{b}})^{2}} \sqrt{\frac{16 \mathrm{Nb}}{1-\mathrm{b}}-1-\frac{(1-2 \sqrt{\mathrm{~b}})^{2}}{(1+\sqrt{\bar{b}})^{2}}+\frac{2(1-2 \sqrt{\mathrm{~b}})}{1+\sqrt{\mathrm{b}}}} . \tag{16}
\end{equation*}
$$

Then, the P.E. of $\left(\frac{q}{p+2 q}\right)^{2}$

$$
\begin{equation*}
=\frac{.6745(1-\sqrt{b})}{4 N(1+\sqrt{\bar{b}})^{2}} \sqrt{\frac{16 \mathrm{Nb}}{1-\mathrm{b}}-1-\frac{(1-2 \sqrt{\mathrm{~b}})^{2}}{(1+\sqrt{\bar{b}})^{2}}+\frac{2(1-2 \sqrt{\mathrm{~b}})}{1+\sqrt{\mathrm{b}}}} . \tag{17}
\end{equation*}
$$

Since the sum of the last three terms under the radical is very small in relation to the first term when N is 100 or more the last three terms may be omitted for approximate calculations.
In this case, P.E. of $\left(\frac{q}{p+2 q}\right)^{2}=\frac{.6745(1-\sqrt{b})}{(1+\sqrt{b})^{2}} \sqrt{\frac{b}{N(1-b)}}$.
I wish to thank Dr. Ralph Tyler for his kind assistance and suggestions in deriving these probable error formulae.

The table follows. It will be seen that values are given for proportions of recessives ranging from .001 to .999 . For extremely rare dominants or recessives, where the proportion of recessives in the population is difficult to determine accurately, other methods of analysis must be used. These will be discussed in a later publication. In the table below,
$B=$ proportion of recessive individuals in the general population,
$\mathrm{R}=$ proportion of recessive offspring to be expected in random matings of dominants with dominants, and
$\mathrm{S}=$ proportion of recessive offspring to be expected in random matings of dominants with recessives.

































a































m 웅 응 ત্ত心












































[^0]:    * Part of the cost of the mathematical composition and the tables in the article is paid by the Galton and Mendel Memortal Fund.

    Genetics 19: 1 Ja 1934

