1. Gerbils and almost all other mammals (including humans) are diploid organisms. That means that they carry two copies of each chromosome (with exception of the sex chromosomes.) Two corresponding versions of the same gene are not necessarily identical. If a gene has several alternate forms, they are called alleles. Different forms of the gene can produce different phenotypic effects. Alleles can be dominant or recessive. If an individual has at least one form of the dominant allele, this allele will determine the phenotype.

In gerbils there is a recessive mutant gene that causes a lethal condition. For the purpose of this problem let the symbol \( A \) denote the normal allele and \( a \) the mutant. In heterozygote individuals who carry both versions of the allele (\( Aa \)), this causes a white spotting color pattern. Homozygotes for the mutant (\( aa \)) die as embryos and are never seen in live gerbils. Since the mutant allele is lethal in the homozygous form, natural selection will occur against the frequency of this allele in a population. The decrease is proportional to the frequency of individuals carrying the gene. Assume that in an isolated population of gerbils in generation \( F_0 \) the frequency of the normal allele is \( p \) and the frequency of the mutant allele is \( q \) (with \( p + q = 1 \)).

(a) Assuming random mating between animals, compute the expected frequencies of offspring with genotypes \( AA \), \( Aa \), and \( aa \), respectively. Use this information to compute the frequencies of the alleles \( A \) and \( a \) in the surviving \( F_1 \) generation in terms of \( p \) and \( q \).

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Fate</th>
</tr>
</thead>
<tbody>
<tr>
<td>( AA )</td>
<td>( p^2 )</td>
<td>Live</td>
</tr>
<tr>
<td>( Aa )</td>
<td>( 2pq )</td>
<td>Live</td>
</tr>
<tr>
<td>( aa )</td>
<td>( q^2 )</td>
<td>Die</td>
</tr>
</tbody>
</table>

Since the homozygous mutant animals will not live, they do not contribute to the \( F_1 \) generation frequency of alleles. Thus, the allelic frequencies in the \( F_1 \) generation are:

<table>
<thead>
<tr>
<th>Allele</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>( \frac{p^2 + 2pq}{p^2 + 2pq + q^2} = \frac{1}{1+q} )</td>
</tr>
<tr>
<td>( a )</td>
<td>( \frac{q^2}{p^2 + 2pq + q^2} = \frac{q}{1+q} )</td>
</tr>
</tbody>
</table>

(b) Write down a general formula for the decrease in frequency of the recessive mutant allele after \( n \) generations.

Hint: It may help, to write down the frequency formulas for the first couple generations, simplify and hunt for a pattern.

<table>
<thead>
<tr>
<th>Generation 0</th>
<th>Generation 1</th>
<th>Generation 2</th>
<th>Generation 3</th>
<th>⋯</th>
<th>Generation ( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p )</td>
<td>( \frac{q}{1+q} )</td>
<td>( \frac{q}{1+\frac{q}{1+q}} ) = ( \frac{q}{1+2q} )</td>
<td>( \frac{q}{1+\frac{q}{1+\frac{q}{1+q}}} ) = ( \frac{q}{1+3q} )</td>
<td>⋯</td>
<td>( \frac{q}{1+\frac{q}{1+\frac{q}{1+\frac{q}{1+\frac{q}{1+\cdots}}}}} ) = ( \frac{q}{1+nq} )</td>
</tr>
</tbody>
</table>

(c) After how many generations can we expect the allele frequency of the recessive mutant to have dropped under 1% of its value in generation \( F_0 \)?

This means to find \( n \), such that

\[
\frac{q}{1+nq} \leq 0.01q \quad \iff \quad q \leq 0.01q + 0.01nq^2
\]
\[
\iffalse
0.99 \leq 0.01nq \implies \frac{99}{q} \leq n
\fi
\]

Thus, the number of years it takes for the lethal recessive trait to drop under 1% is reciprocal to its original frequency in the population (rare traits take a long time to vanish).

2. In a population of parent pea plants, the frequency of the red-flower color allele (R) is 0.7 and the frequency of the white flower color allele (r) is 0.3. The red allele (R) is dominant to the white allele (r). Assume that two alleles are passed down independently from parent plants to their offspring. In other words, the probability that a randomly chosen parent plant will pass an R-allele down to its offspring is 0.7 and the probability that the plant will pass down an r-allele is 0.3.

(a) What is the probability that an offspring plant has genotype Rr?

A plant with genotype Rr may have inherited R from “mom” (with probability 0.7) and r from “dad” (with probability 0.3) or vice versa! The probability of inheriting Rr is thus

\[2 \cdot 0.7 \cdot 0.3 = 0.42\]

(b) What is the probability that an offspring plant has red flowers?

Since the red color allele R is dominant, plants with genotype RR and Rr will have red flowers. The probability that a plant has either RR or Rr genotype is

\[P(\text{RR or Rr}) = 0.7^2 + 2 \cdot 0.7 \cdot 0.3 = 0.91\]

(c) What is the probability that a plant with red flowers in the parent generation will pass on an R allele to its offspring?

A red flowered plant has either genotype RR or genotype Rr. These two events are mutually exclusive. Hence the probability that a plant with red flowers is of genotype RR is

\[P(\text{RR} \mid \text{red}) = \frac{0.49}{0.49 + 0.42} = 0.538\]

while the probability that a red flowered plant has genotype Rr is

\[P(\text{Rr} \mid \text{red}) = \frac{0.42}{0.49 + 0.42} = 0.462.\]

Hence the probability that a red flowered plant passes on an R allele

\[P(\text{pass on R} \mid \text{red}) = P(\text{pass on R} | \text{Rr})P(\text{Rr}) + P(\text{pass on R} | \text{RR})P(\text{RR})\]

\[= \frac{1}{2} \cdot 0.462 + 1 \cdot 0.538 = 0.769\]

(d) A parent pea plant with red flowers and known genotype Rr is bred to another parent pea plant with red flowers but unknown genotype. What is the probability that these two plants will have an offspring of genotype RR (Rr, rr)?
Let’s call the Rr genotype parent “mom” and the parent with unknown genotype “dad”. Since “dad” has red flowers, he may be either of genotype RR (with probability 0.538) or Rr (with probability 0.462). If “dad” is of genotype RR he will pass an R allele with probability 1 and if he is of Rr genotype he will pass either allele with probability $\frac{1}{2}$.

Thus,

$$P(\text{offspring is RR})$$

$$= P(\text{R from mom and R from dad}|\text{dad is RR})P(\text{dad is RR})$$

$$+ P(\text{R from mom and R from dad}|\text{dad is Rr})P(\text{dad is Rr})$$

$$= \frac{1}{2} \cdot 1 \cdot 0.538 + \frac{1}{2} \cdot \frac{1}{2} \cdot 0.462 = 0.3845.$$

Similarly,

$$P(\text{offspring is Rr})$$

$$= P(\text{R (mom) and r (dad) or r (mom) and R (dad)}|\text{dad is RR})P(\text{dad is RR})$$

$$+ P(\text{R (mom) and r (dad) or r (mom) and R (dad)}|\text{dad is Rr})P(\text{dad is Rr})$$

$$= \left(\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1\right) \cdot 0.538 + \left(\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2}\right) \cdot 0.462 = 0.5.$$

And,

$$P(\text{offspring is rr})$$

$$= P(\text{r (mom) and r (dad)}|\text{dad is RR})P(\text{dad is RR})$$

$$+ P(\text{r (mom) and r (dad)}|\text{dad is Rr})P(\text{dad is Rr})$$

$$= \frac{1}{2} \cdot 0 \cdot 0.538 + \frac{1}{2} \cdot \frac{1}{2} \cdot 0.462 = 0.1155.$$

Note that all three of these probabilities must sum to one.

3. R can be used to generate sequences of random numbers. For instance, the command `rnorm(10, µ, σ)` (with some numbers substituted for µ and σ) will generate ten random numbers from a Normal distribution with mean µ and standard deviation σ. Find ten random numbers generate by R below:

$$3.28 \ 3.52 \ 2.25 \ 2.33 \ 4.29 \ 3.82 \ 3.89 \ 4.43 \ 1.91 \ 2.28$$

For your convenience, the same data are available in the file “RandomNumbers.txt” on the homework website.

(a) Use a one-sample $t$-test and significance level $\alpha = 0.05$ to test the hypothesis that the mean $\mu$ used to generate those random numbers was $\mu = 2.5$. Show the computations you do to carry out the test (i.e., do not use R to carry out the test for you). To obtain the $p$-value, you may use your calculator, a table from a textbook or the R-command `pt()` (p stands for probability and t stands for $t$-distribution).
First, we need to calculate the sample mean \( \bar{x} \) and the sample standard deviation \( s \).

\[
\bar{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = 3.2,
\]

\[
s = \sqrt{\frac{1}{9} \sum_{i=1}^{10} (x_i - \bar{x})^2} = 0.933
\]

Then the value of the test statistic becomes

\[
t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{3.2 - 2.5}{0.933/\sqrt{10}} = 2.373
\]

Since the alternative is unspecified, we’ll use a two-sided test (\( H_a : \mu \neq 2.5 \)). Thus, the \( p \)-value for this test is

\[
P(|t_{df=9}| \geq 2.373) = 2 \cdot P(t_{df=9} > 2.373) = 0.0417
\]

Since this \( p \)-value is smaller than the significance level \( \alpha \) we reject the null hypothesis and conclude that the random numbers were not generated with \( \mu = 2.5 \).

(b) Suppose that I would use \( \mu = 2.5 \) and \( \sigma = 1 \) to generate many independent sets of ten random numbers, each. If I were to repeatedly carry out the same test you did in (a) for each new vector of ten numbers, how often, on average, would you expect to reject the null hypothesis? Explain how you come to your conclusion.

In other words, the question is asking: if the null hypothesis were true, how often would the \( t \)-test reject in the long run. Any test rejects when its \( p \)-value is less than the significance level \( \alpha \). The probability that that happens (if the null hypothesis is correct) is \( \alpha \) itself. Thus, in the long-run, on average 5% of tests thus conducted would reject the null hypothesis (even though it is true).

(c) Read the data into R from the file or directly type the data into R. Use the \texttt{t.test()} command to carry out the same \( t \)-test you performed in (a). Cut and paste the resulting output to hand in with your homework.

\[
> \text{t.test(x, mu=2.5, alternative="two.sided")}
\]

\begin{verbatim}
One Sample t-test

data:  x
t = 2.3717, df = 9, p-value = 0.04179
alternative hypothesis: true mean is not equal to 2.5
95 percent confidence interval:
2.532325 3.867675
sample estimates:
mean of x
3.2
\end{verbatim}

The reason that the \( p \)-value computed by R does not exactly match ours is the fact that our sample standard deviation estimate was rounded.