

### **3. ANSWERS TO PROBLEMS**

#### **1. Generate a random sample of 50 rat families with litter sizes of $n=10$ and where the probability of any one offspring being male is $p=0.5$ .**

a) Individual student's graphs will vary, but the histograms and boxplots should be roughly symmetric, with perhaps some similarity to a normal (bell-shaped) distribution.

b) The population mean and standard deviation is determined by the parameters of the binomial distribution you provided:

$$\text{mean } \mu = n p = (10)(0.5) = 5$$

$$\text{standard deviation } \sigma = (npq)^{1/2} = (10)(0.5)(0.5)^{1/2} = 1.581$$

The sample mean and standard deviation will be similar, but not identical, to these parameter values because of the noise inherent in random sampling.

c) Results of individual students will vary, but the new histogram should be similar, but not identical, to the previous histogram. The change may be attributed to the noise inherent in random sampling.

d) The binomial distribution with  $p=0.90$  will have a different shape from that when  $p=0.5$ . There will be more males on average (higher mean) and a reduced spread in the number of males (lower variance). The new histogram should be skewed to the left. In the new population:

$$\text{mean } \mu = n p = (10)(0.9) = 9$$

$$\text{standard deviation } \sigma = (npq)^{1/2} = (10)(0.9)(0.1)^{1/2} = 0.949$$

e) Increasing sample size should have little effect on the shape of the distribution on average, but the distribution will conform more closely to the population distribution. There will be less variation between the histograms produced by different individuals (and between separate random samples obtained when the "Apply" button is clicked again and again. With 1000 values the sample mean and standard deviation should lie closer on average to the corresponding population parameters than was the case with the smaller sample of  $n=50$ . The estimates based on the larger sample are therefore more reliable.

f) The distribution should become more symmetric and "bell-shaped" (normal) with larger  $n$ .

#### **2. A Poisson distribution of Asian Gypsy moths captured in traps, mean is 0.5.**

a) The Poisson distribution with mean of 0.5 should be positively skewed. The number of individuals trapped can take on only integer values. The mean should be similar, but not identical, to 0.5. The standard deviation should be around 0.707, which is the square root of 0.5.

- b) As the mean of the Poisson distribution is reduced, the shape of the distribution should tend to become more right-skewed.
- c) When the mean of the Poisson distribution is increased, the shape of the distribution becomes increasingly symmetric and eventually comes to resemble the normal (bell-shaped) distribution.
- d) The variance should be approximately the same as the mean. In each population, mean and variance are identical (standard deviation and square root of the mean are identical) when the variable has a Poisson distribution.

### 3. Binomial test of changes in butterfly range limits

- a)  $H_0$ : Northward changes and southward changes in northern limits are equally frequent ( $p = 0.5$ ).  
 $H_A$ : Northward changes and southward changes in northern limits are not equally frequent ( $p \neq 0.5$ ).
- b)  $P = \text{Prob}(\text{result at least as extreme as that observed when } H_0 \text{ is true})$   
 $= 2 \times [\text{Prob}(X=18) + \text{Prob}(X=19) + \text{Prob}(X=20) + \text{Prob}(X=21)]$   
 $= 0.0015$

**Note:** I have used a two-tailed test. That is, my alternative hypothesis is that northern limits changed in one direction (northward or southward) more frequently than in the opposite direction. However, it might be argued that a one-tailed test is more appropriate here. Perhaps our alternate hypothesis should be that northward expansion of northern range limits was more frequent than southward retraction, because the point of the study was to test effects of global warming (which would predict northward expansion of northern limits, not southward retraction). In this case, our  $P$ -value would be half of the value calculated above, since we would leave the “ $2 \times$ ” out of the calculation of  $P$ . We will have further opportunity to debate the merits of one- and two-tailed tests. In general, a one-tailed test is justified only when there are solid prior reasons to believe that one tail of possibilities under the alternative hypothesis is unlikely or inconceivable.

### 4. Distribution of new mutations.

- a)  $\bar{X} = 0.883, s^2 = 0.958$
- b) The number of new mutations should follow a Poisson distribution.
- c) Use the formula for the Poisson distribution (with a mean  $\mu=0.8833$ ) to calculate the probability  $P(X)$  that an individual has  $X = 0, 1, 2, 3,$  and  $4$  new mutations. You can use the **Poisson Probability** function of the calculator or do it step by step using the Poisson formula. Here are the numbers I obtained:

<b>Number of new mutations</b>	<b>Observed number of individuals</b>	<b>Expected number of individuals</b>
0	25	24.80
1	22	21.91
2	9	9.68
3	3	2.85
4	1	0.63
>4	0	0.13
<b>Total</b>	<b>60</b>	<b>60.00</b>

These are quite close to the observed frequencies, suggesting a good fit to the Poisson distribution. Formal tests of the goodness-of-fit will be covered later in a later lab.