SOLUTIONS TO LAB 2 ASSIGNMENT

Question 1

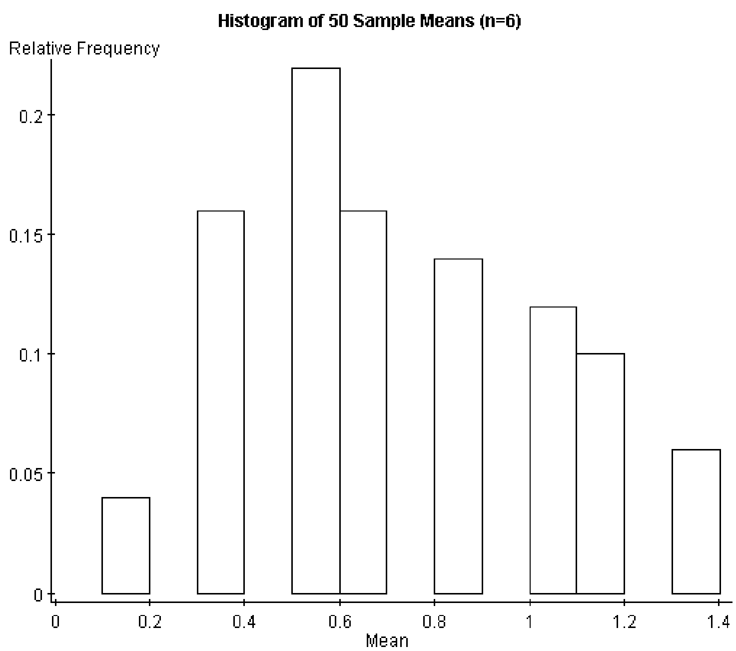
(a) Using the calculator tool, we can see that the distribution of the number of flaws (the Poisson distribution) is highly skewed to the right. The distribution has a single peak at zero. The pattern indicates that most observations are at zero or one, or that the typical number of flaws per sheet is either 0 or 1.

(b) The probability that there are no flaws is $P(X = 0) = 0.4966$ (to 4 decimal places). About 50% of glass sheets are in perfect condition.

(c) Assuming independence between glass sheets, the probability that all glass sheets in a randomly selected box (of 6 glass sheets) are in perfect condition is equal to $(0.4966)^6 = 0.0150$ (to 4 decimal places). Thus, about 1.5% of boxes contain 6 glass sheets in perfect condition.

Question 2

(a) First, it is necessary to obtain the sample means of the 50 samples of 6 observations (*lab2a.txt*) with the *Summary Statistics (Column)* feature (mean should be the only statistic selected). Moreover, the "Store output in the data table" check box should be checked to create the column "*Mean*" of the 50 sample means in the data table. The column "*Mean*" of the 50 averages can now be used to obtain a histogram of the sample means. The histogram of the 50 sample means based on the 50 random samples of 6 sheets with bins starting at 0 and a width of 0.1 is shown below:



(b) As you can see, the histogram of averages based on 50 random samples of size 6 is skewed to the right. The right skewness is inherited from the parent distribution, which is highly skewed to the right. However, the degree of right skewness is much smaller for the distribution of averages than for the parent distribution. The centers of both distributions are very similar (between 0 and 1), but their spreads are very different. The spread of observations in the distribution of averages is quite smaller than the spread of observations in the parent distribution (comparing the ranges).

(c) The mean, standard deviation and standard error of the 50 averages can be calculated with the *Summary Statistics (Column)* feature in the *Stat* menu. The results are displayed below:

Summary statistics:

Column	**Mean**	**Std. Dev.**	**Std. Err.**	**n**
Mean	0.71	0.32087255	0.04537823	50

The mean of the 50 sample means is 0.71 and the standard deviation of the 50 sample means is 0.3209 (to 4 decimal places). In general, the standard deviation here is an estimate of the standard deviation of the sampling distribution of a sample mean. In our case, the sample mean is the mean of 50 sample means and the standard deviation of 0.3209 reported in the output measures the variability of the 50 sample means. Each of the 50 random samples consists of 6 observations from the population. The standard error is 0.04538.

According to the properties of the sampling distribution of a sample mean, if a population has a mean μ and standard deviation σ, then the distribution of the sample mean has a mean and standard deviation defined by the formulas:

$$\mu_{\bar{X}} = \mu_X , \ \sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}$$

For a Poisson random variable, X, with parameter $\lambda = 0.7$,

$$\mu_X = \sigma_X^2 = \lambda = 0.7$$

Thus, the mean of the 50 averages should be reasonably close to the value of 0.7 and the standard deviation should be close to $\sqrt{0.7}/\sqrt{6} = 0.341565$. The observed values of 0.71 and 0.3209 are reasonably close to the predicted values.

Note: Information below is additional but not required for lab.
The standard deviation of the 50 sample means (each based on a random sample of 6 observations from the distribution) is

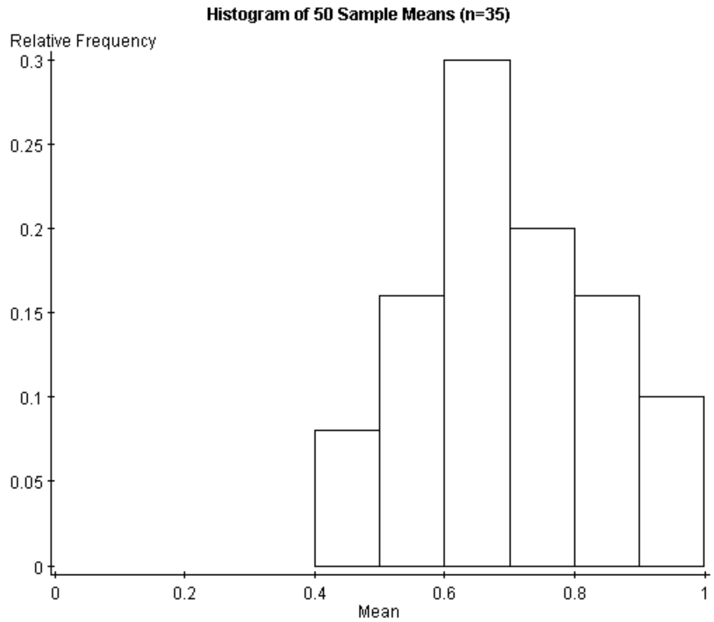
$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{0.341565}{\sqrt{50}} = 0.048305$$

The value of the standard error 0.04537 reported in the above output is reasonably close to the above value. Notice that we cannot expect the predicted values to be identical to the corresponding observed values because the two observed values are based on a sample of averages but not on all possible averages.

Question 3

In this part, it is necessary to calculate the average for each of the 50 samples of 35 observations (file *lab2b.txt*).

(a) The histogram of the 50 sample means is displayed below:



(b) As you can see, the histogram of the 50 sample means (from samples of 35) is approximately bell-shaped, which is associated with the normal curve. The center of the distribution is close to 0.7 and the spread is much smaller than the spread of the parent distribution. The spread is also smaller than the spread for the distribution of sample means based on samples of size 6 obtained in Question 2, part (a).

(c) The observed mean, standard deviation and standard error are:

Summary statistics:

Column	n	Mean	Std. Dev.	Std. Err.
Mean	50	0.69371426	0.13985653	0.0197787

The mean and standard deviation of the 50 sample means based on samples of 35 are 0.6937 and 0.1399, respectively. The mean is roughly the same as it was based on the samples of size 6, but the standard deviation is much smaller. The standard error is 0.01978.

The mean of the 50 averages (0.6937) is reasonably close to the value of 0.7 that it should be under its theoretical distribution. The standard deviation for the 50 averages (0.1399) is also quite close to the value predicted by the theory for all possible samples of 50 observations: $\sigma_X / \sqrt{n} = \sqrt{0.7} / \sqrt{35} = 0.141421$

Notice that, we cannot expect the predicted values to be identical to the corresponding observed values because the two observed values are based on a sample of averages but not on all possible averages.

Question 4

(a) Based on Question 1, part (c), the fraction of boxes (samples) with six flawless sheets is equal to 0.015 (assuming independence, which is reasonable given the large production run). Thus, we expect $50*0.015 \approx 0.75$ boxes containing six flawless sheets. Since we can't have 0.75 boxes, we'd probably expect 0 or 1 boxes. The observed number of boxes is 0.

(b) The *Calculators (Normal)* feature in the *Stat* menu calls for the mean and standard deviation of the sampling distribution of a sample mean. For samples of 35, the two values are equal to 0.7000 and 0.1414 (to four decimal places), respectively. Then, the probability that the mean will be equal to or exceed 0.90 is 0.0786 (to four decimal places).

The calculation is based on the assumption that the probability distribution of a sample mean approximately follows a normal distribution. Given the large sample size of $n = 35$ ($n \geq 30$), the assumption of normality is justified. There are actually five samples in the data table that produce the average of 0.90 or higher, or an observed fraction of 5/50 = 0.1000. Accounting for some variability, the probability of 0.0786 is consistent with the observed relative frequency.

LAB 2 ASSIGNMENT MARKING SCHEMA

Proper Header and Appearance: 10 points

Question 1 (10)

(a) Description of the shape of the distribution: 3 points
Evaluation of the typical number of flaws: 2 points
(b) Probability that there are no flaws in a randomly selected sheet: 3 points
(c) Fraction of boxes with no defective glass sheets: 2 points

Question 2 (20)

(a) Histogram: 6 points
(b) Description of the shape of the histogram: 3 points
Comparison with the parent distribution: 3 points
(c) Mean, standard deviation, and standard error: 3 points
Standard deviation: 2 points
Comparison with the values predicted by the theory: 3 points

Question 3 (25)

(a) Histogram: 6 points
(b) Description of the shape of the histogram: 3 points
Comparison with the parent distribution: 3 points
Comparison with the histogram in Question 2: 3 points
(c) Mean, standard deviation, and standard error: 3 points
Comparison with the values obtained in Question 2: 3 points
Comparison with the values predicted by the theory: 2 points
Conclusion: 2 points

Question 4 (10)

(a) Observed count of boxes of 6 with flawless sheets: 2 points
Comparison with the probability: 2 points
(b) Probability: 3 points
Observed fraction for samples of size 35: 2 points
Comparison: 1 point

TOTAL = 10 + 10 + 20 + 25 + 10 = 75