## B.3 Homework

*This worksheet accompanies the article found at http://www.amstat.org/publications/jse/v10n3/haller.html*

**1.** Use simulation to obtain the values of sample proportions of base landings for 100 different samples of sizes $n$ = 5, 15, 40, 80 and 120, where the true proportion of base landings is assumed to be $p$ = 0.30. For each sample size, construct a histogram of the 100 simulated sample proportions. Do this using a computer software package. Describe the distribution of the sample proportion values for each of the sample sizes.

**2.** Use simulation to obtain the values of sample proportions of base landings for 15 different samples for true proportions of base landings: $p$ = 0.10, 0.25, 0.50, 0.75 and 0.90 all with a sample size of $n$ = 100. Record the proportions as decimals with two significant digits. Do this using a computer software package.

**Table 1.** Simulated proportions of base landings for HERSHEY'S KISSES.

| Repetition | $p$ = 0.10 | $p$ = 0.25 | $p$ = 0.50 | $p$ = 0.75 | $p$ = 0.90 |
|---|---|---|---|---|---|
| 1 | | | | | |
| 2 | | | | | |
| 3 | | | | | |
| 4 | | | | | |
| 5 | | | | | |
| 6 | | | | | |
| 7 | | | | | |
| 8 | | | | | |
| 9 | | | | | |
| 10 | | | | | |
| 11 | | | | | |
| 12 | | | | | |
| 13 | | | | | |
| 14 | | | | | |
| 15 | | | | | |

**3.** Calculate the variance of each column from the table above and put the value in the table below. Record the variances as decimals with four significant digits.

**Table 2.** Variance of the sample proportion values.

| $p$ | $p = 0.10$ | $p = 0.25$ | $p = 0.50$ | $p = 0.75$ | $p = 0.90$ |
|---|---|---|---|---|---|
| Variance | | | | | |

**4.** Use your calculations from Question 3 to complete the following.

(a) Construct a scatter plot of the variance of the estimated proportion (vertical axis) versus the true proportion (horizontal axis).

(b) For which value of $p$ is the variance the highest? For which values of $p$ is the variance the smallest?

**5.** The table below has the results for an experiment where $n = 100$ and $p = 0.50$.

**Table 3.** Sample proportions obtained from an experiment with $n = 100$ and $p = 0.50$.

| Trial | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\hat{p}$ | .53 | .50 | .56 | .39 | .61 | .52 | .46 | .53 | .49 | .46 | .52 | .51 | .53 | .51 | .49 |

(a) Calculate the mean of all fifteen sample proportions and the standard deviation of all fifteen sample proportions.

(b) Calculate the first quartile (Q1), the median (M) and the third quartile (Q3) of the fifteen sample proportions. Recall that these values are the 25th percentile, the 50th percentile and the 75th percentile.

(c) Using your knowledge of calculating Q1 and Q3, calculate the 12.5th percentile, 37.5th percentile, 62.5th percentile and the 87.5th percentile of the fifteen sample proportions. You now have the empirical percentiles.

(d) Calculate the 12.5th percentile, 25th percentile, 37.5th percentile, 50th percentile, 62.5th percentile, 75th percentile, and the 87.5th percentile for the normal distribution with a mean and standard deviation from part (a) above.

Record your results from parts (b), (c), and (d) in Table 4.

**Table 4.** Empirical and Normal Percentiles with $n = 100$ and $p = 0.50$.

| Percentile | 12.5th | 25th | 37.5th | 50th | 62.5th | 75th | 87.5th |
|---|---|---|---|---|---|---|---|
| Empirical | | | | | | | |
| Normal | | | | | | | |

(e) Construct a scatter plot of the empirical percentiles versus the normal percentiles. Does the scatter plot seem to follow a straight line? If so, then the sample proportions can be adequately modeled with a normal distribution.

**6.** The table below has the results for an experiment where $n = 100$ and $p = .04$.

**Table 5.** Sample proportions obtained from an experiment with $n = 100$ and $p = .04$.

| Trial | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\hat{p}$ | .02 | .05 | .03 | .06 | .03 | .02 | .05 | .04 | .03 | .08 | .03 | .04 | .03 | .05 | .02 |

(a) Calculate the mean of all fifteen sample proportions and the standard deviation of all fifteen sample proportions.

(b) Calculate the first quartile (Q1), the median (M) and the third quartile (Q3) of the fifteen sample proportions. Recall that these values are the 25th percentile, the 50th percentile and the 75th percentile.

(c) Using your knowledge of calculating Q1 and Q3, calculate the 12.5th percentile, 37.5th percentile, 62.5th percentile and the 87.5th percentile of the fifteen sample proportions. You now have the empirical percentiles.

(d) Calculate the 12.5th percentile, 25th percentile, 37.5th percentile, 50th percentile, 62.5th percentile, 75th percentile, and the 87.5th percentile for the normal distribution with a mean and standard deviation from Part (a) above.

Record your results from parts (b), (c), and (d) in Table 6.

**Table 6.** Empirical and Normal Percentiles with $n = 100$ and $p = .04$.

| Percentile | 12.5th | 25th | 37.5th | 50th | 62.5th | 75th | 87.5th |
|---|---|---|---|---|---|---|---|
| Empirical | | | | | | | |
| Normal | | | | | | | |

(e) Construct a scatter plot of the empirical percentiles versus the normal percentiles. Does the scatter plot seem to follow a straight line? If so, then the sample proportions can be adequately modeled with a normal distribution.

**7.** Use simulation to obtain the values of sample proportions for 100 different samples of size $n = 100$, where the true proportion is assumed to be $p = .01$. Construct a histogram of the 100 simulated sample proportions. Do this using a computer software package.

(a) Your histogram should look positively skewed (right skewed) indicating that the sample proportion values do not have a normal distribution. Explain why you think this happens.

(b) How would a histogram of 100 simulated sample proportions for $n = 100$ and $p = 0.99$ compare to a histogram of 100 simulated sample proportions for $n = 100$ and $p = 0.01$?