Assignment #5

Due on Wednesday November 5, 2008

Read Chapter 4 in the class notes, *Introduction to Estimation,* in the course webpage at `http://pages.pomona.edu/~ajr04747`

Read Chapter 5 on *Sampling Distributions* in Moore, McCabe and Craig.

Read Section 6.1 on *Estimating with Confidence* in Moore, McCabe and Craig.

Background Information.
To estimate binomial probabilities, use the R function `dbinom()`. For instance, if $Y \sim B(n,p)$, then $p_Y(k) \approx \text{dbinom(k,n,p)}$.

The function `pbinom()` gives the cumulative distribution function for $Y \sim B(n,p)$ as follows

$$P(Y \leqslant k) \approx \text{pbinom(k,n,p)}.$$

The cumulative distribution function for a $N(\mu, \sigma^2)$ random variable, X, may be estimated by the `pnorm()` function as follows

$$F_X(x) \approx \texttt{pnorm}(x, \mu, \sigma).$$

Note that we input the standard deviation, σ, and not variance, σ^2. To approximate normal probabilities in R you may use

$$P(a < X \leqslant b) \approx \text{pnorm(b,n,p)} - \text{pnorm(a,n,p)}.$$

The values of the density function, f_X, for $X \sim N(\mu, \sigma^2)$ can be estimated in R using the `dnorm()` function as follows

$$f_X(x) \approx \text{dnorm}(\text{x}, \mu, \sigma) \quad \text{for all } x \in \mathbf{R}.$$

Do the following problems

1. (Typographical Errors[1]) Typographical and spelling errors can be either "nonword errors" or "word errors." A nonword error is not a real word, as when "the" is typed as "teh." A word error is a real word, but not the right word, as when "lose" is typed as "loose." Spell–checking software will catch nonword errors but not word errors. Human proofreaders catch around 70% of word errors. You ask a fellow student to proofread an essay in which you have deliberately made 10 word errors.

[1] Adapted from Exercise 5.13 in Moore, McCabe abd Graig, *Introduction to the Practice of Statistics,* Sixth Edition, pp. 331–332

(a) If the student matches the usual 70% rate, what is the distribution of the number of errors caught? What is the distribution of the number of errors missed?

(b) Missing 4 or more of the 10 errors seems a poor performance. What is the probability that a proofreader who catches 70% of the word errors misses 4 or more out of 10?

2. (Typographical Errors (continued)[2])

(a) What is the mean number of errors caught? What is the mean number of errors missed?

(b) What is the standard deviation, σ, of the number of errors caught?

(c) Suppose that a proof reader catches 90% of word errors, so that $p = 0.9$. What is the standard deviation, σ, in this case? What is σ is $p = 0.99$? What happens to the standard deviation as p approaches 1?

3. Let X_i denote the number on the face that comes up in the i^{th} roll of a balanced die.

(a) Give the expected value and variance for each X_i, for $i = 1, 2, 3, \ldots$

(b) n rolls of the die constitutes a random sample of size n:

$$X_1, X_2, X_3, \ldots, X_n.$$

Compute the sample mean

$$\overline{X}_n = \frac{X_1 + X_n + \cdots + X_n}{n},$$

and determine its expected value and variance.

4. *(Continuation of Problem 3)*

(a) Use R to simulate rolling the die $n = 100$ times. This simulates collecting a random sample

$$X_1, X_2, X_3, \ldots, X_n$$

of size $n = 100$. Perform 1000 repetitions of the experiment to generate a simulation of the sampling distribution of $\overline{X}_n$. Plot a histogram of the simulations.

[2]Adapted from Exercise 5.15 in Moore, McCabe abd Graig, *Introduction to the Practice of Statistics,* Sixth Edition, p. 332

(b) Use the histogram generated in the previous part to obtain the probability distribution for the simulations of $\overline{X}_n$. In the same graph, plot the normal curve that approximates the distribution of $\overline{X}_n$ according to the Central Limit Theorem.

5. (Fuel Efficiency[3]) Computers in some vehicles calculate various quantities related to performance. One of those quantities is fuel efficiency, or gas mileage, usually expressed as miles per gallon (mpg). For one vehicle equipped in this way, the mpg were recorded each time the gas tank was filled, and the computer was then reset. The mpg values are contained in the MS Excel file `FuelEfficiencyData.xls`, which may be downloaded from

   `http://pages.pomona.edu/~ajr04747`.

   Assume that the standard deviation, σ, for the mpg random variable is known to be 3.5 mpg.

   (a) Give the variance of the sample mean for a random sample of size n; that is, compute $\mathrm{Var}(\overline{X}_n)$.

   (b) Give the 95% confidence interval for the true mean mileage of the vehicle.

[3] Adapted from Exercise 6.26 in Moore, McCabe abd Graig, *Introduction to the Practice of Statistics,* Sixth Edition, p. 371