1. Recall the following problem from test 1:

   Infection with an apparently harmless, newly recognized virus called hepatitis G, seems to interfere with HIV, slowing its progress and prolonging survival of AIDS patients, according to an Associated Press article in the September 6, 2001 edition of the Boston Globe. A study by researchers at the Iowa City Veterans Affairs Medical Center and the University of Iowa looked at 362 HIV-infected patients treated between 1988 and 1999. The researchers found that 144 of these patients were also infected with hepatitis G. Of the 144 who were infected with hepatitis G, 41 died during four years of follow-up compared with 123 of the 218 who were not infected with hepatitis G.

   Assume the 144 patients with both HIV and hepatitis G form a random sample from the population of all such individuals, and that the second sample of 218 patients form a random sample from the population of all patients with HIV but no hepatitis G.

   (a) **(10 points)** Obtain a point estimate of the difference between the proportion of the population of patients infected with both HIV and hepatitis G who survive four years and the proportion of the population of patients infected with HIV but not with hepatitis G who survive four years.

   ANS: *The point estimate is* $\hat{p}_1 - \hat{p}_2 = 103/144 - 95/218 = 0.2795$

   (b) **(10 points)** Obtain a 95% approximate score confidence interval for the difference estimated in (a).

   ANS: *Since* $z_{0.975} = 1.96$,

   $$\tilde{p}_1 = \frac{103 + 0.25 \cdot 1.96^2}{144 + 0.5 \cdot 1.96^2} = 0.7124, \ \tilde{p}_2 = \frac{95 + 0.25 \cdot 1.96^2}{218 + 0.5 \cdot 1.96^2} = 0.4363$$

   The interval is then

   $$0.7124 - 0.4363 \pm 1.96\sqrt{\frac{0.7124(1 - 0.7124)}{144} + \frac{0.4363(1 - 0.4363)}{218}} = (0.1771, 0.3751)$$

   (c) **(10 points)** Explain what "95% confidence" means here.

   ANS: *Approximately 95% of all such intervals obtained in repeated sampling from the two populations will contain the true difference in population proportions* $p_1 - p_2$.

   (d) **(10 points)** Explain in layman's terms what the interval you computed in (a) means in terms of the population proportions involved.

   ANS: *We estimate that the percentage of the population of HIV patients infected with hepatitis G that survives during four years of follow-up exceeds the percentage of HIV patients not infected with hepatitis G that survives during four years of follow-up by somewhere between 17.71 and 37.51%.*

2. A computer manufacturer randomly samples computer mice from the production line and subjects them to wear tests. In one such test, the left mouse button is pressed repeatedly by a testing machine until it fails. Data on the number of press-release cycles until failure in 12 test mice has mean 6709.6 and standard deviation 117.2. The process is stationary, and the data show no evidence of nonnormality.

   (a) **(10 points)** Construct a level 0.99 confidence interval for the population mean cycles to failure.

   ANS: *The interval is*

   $$\overline{y} \pm \frac{s}{\sqrt{n}} t_{11,0.995} = 6709.6 \pm \frac{117.2}{\sqrt{12}} 3.1058 = (6604.5, 6814.7)$$

   (b) **(20 points)** Construct an interval based on the 12 values in the data set which you are 99 percent confident will contain the number of cycles to failure of the next mouse tested. If the next mouse fails after 6208 cycles, what will you conclude based on the interval you calculated?

ANS: *We construct a level 0.99 prediction interval. The interval is*

$$\overline{y} \pm s\sqrt{1+\frac{1}{n}}t_{11,0.995} = 6709.6 \pm 117.2\sqrt{1+\frac{1}{12}}3.1058 = (6330.7, 7088.5)$$

Since 6208 is below the interval, we conclude that the performance of this mouse is inferior to that of the population as represented by the 12 mice in the sample.

(c) **(20 points)** Construct an interval you are 99% confident will contain at least 95% of all cycles to failure values in the population. Interpret this interval in terms of "99% confidence".

ANS: *We construct a level 0.99 tolerance interval for proportion 0.95 of the population values. The interval is*

$$\overline{y} \pm K \cdot s = 6709.6 \pm 3.870 \cdot 117.2 = (6256.0, 7163.2)$$

Interpretation: In repeated sampling, approximately 99% of all intervals constructed will contain at least 95% of all population values.

3. **(10 points)** Consider a level L confidence interval for the mean μ in the C+E model where σ is known. If I want to double the precision of the interval (i.e., make it half as wide), how much must I increase the sample size? How about if I want to triple the precision? Give a general formula for how much I must increase the sample size n to make the interval k times as precise.

ANS: *We know from formula (5.20), p. 257 of the text, that the sample size is inversely proportional to the square of the desired precision. Therefore, to double the precision, we need to quadruple the sample size. To triple the precision, we need 9 times the sample size. To increase precision by a factor of k, we need to obtain k^2 times the number of observations.*