ANSWERS TO TEST NUMBER 4

Question 1: (20 points)

Given the distribution of the population

x:	0	1	2
$P(x)$:	.10	.15	.75

the probabilities of obtaining each of the nine possible sample combinations can be calculated as follows:

X_1	X_2	$\bar{X}$	$(P(X_1))(P(X_2))$
0	0	0	$(.10)(.10) = .01$
0	1	.5	$(.10)(.15) = .015$
0	2	1	$(.10)(.75) = .075$
1	0	.5	$(.15)(.10) = .015$
1	1	1	$(.15)(.15) = .0225$
1	2	1.5	$(.15)(.75) = .1125$
2	0	1	$(.75)(.10) = .075$
2	1	1.5	$(.75)(.15) = .1125$
2	2	2	$(.75)(.75) = .5625$

The probability that $\bar{X}$ could take each of its 5 possible values can then be obtained by adding together the probabilities of obtaining the individual samples that result in each of those values.

$\bar{X}$	$P(\bar{X})$	
0	.01	= .01
.5	.015 + .015	= .03
1	.075 + .0225 + .075	= .1725
1.5	.1125 + .1125	= .225
2	.5626	= .5625

[Give yourself 8 points for correctly setting up the calculations of the possible sample means and 8 points for correctly calculating the sampling distribution of the mean.]

The population mean is

$$E\{X\} = \sum_{1}^{3} X_i P_{X_i} = (0)(.10) + (1)(.15) + (2)(.75) = 1.65$$

and the population variance is

$$\begin{aligned}
\sigma^2\{X\} &= \sum_{1}^{3} (X_i - E\{X\})^2 P_{X_i} \\
&= (0 - 1.65)^2(.10) + (1 - 1.65)^2(.15) + (2 - 1.65)^2(.75) \\
&= (2.7225)(.10) + (.4225)(.15) + (.1225)(.75) \\
&= .27225 + .063375 + .091875 \\
&= .4275
\end{aligned}$$

The variance of the sample mean is therefore

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$$

which yields

$$\sigma = \sqrt{.4275/10} = .206761$$

for $n = 10$ and

$$\sigma = \sqrt{.4275/100} = .065383$$

for $n = 100$.

[Give yourself one point for knowing how to correctly calculate each of the the population mean, the population variance, and the two sample standard deviations (four points in total).]

Question 2: (15 points)

False. It is the probability distribution of the sample mean that approaches normality as the sample size increases.

[Give yourself 15 points for demonstrating this understanding of what is meant by the central limit theorem.]

Question 3: (15 points)

a) An estimator is a statistic used to estimate a parameter value—for example, the sample mean is a statistic that is often used as an estimator of the population mean.

b) An estimate is the value of the estimator for a particular sample.

c) Unbiasedness a property of estimators that occurs when the expectation of the estimator equals the parameter being estimated. That is, an estimator S is an unbiased estimator of a parameter θ if

$$E\{S\} = \theta$$

The bias of an estimator is equal to $E\{S\} - \theta$.

d) Consistency is a property of estimators that occurs when increasing the sample size results in estimates that tend to be closer to the parameter being estimated.

e) Efficiency us a property of unbiased estimators. Of two unbiased estimators, the one with the smallest variance is the more efficient.

[Give yourself 3 points for demonstrating your understanding of each of the five concepts.]

Question 4: (25 points)

a) The expected value of the mean return from a sample of 14 from the group of stocks is the population mean—10% per year. The variance is the population variance divided by the sample size—$\sigma^2/n = 4/14 = .285714$. The standard deviation is then .534522.

b) Because the sample size is small we use the t-distribution. The value of $t(1-\alpha/2; n-1) = t(.95; 13)$—i.e., the value of t to the left of which the probability mass is .95—is 1.7709. You can find this up by firing up XLISPSTAT and entering the following command (the line below it is the program's response):

```
> (t-quant .95 13)
1.70933395265046
```

Thus we have an upper confidence limit of

$$U = 10 + t(.95; 13)\,\sigma/\sqrt{n} = 10 + (1.7709)(.534522) = 10 + .946603 = 10.946603$$

and a lower confidence limit of

$$L = 10 - t(.95; 13)\,\sigma/\sqrt{n} = 10 - (1.7709)(.534522) = 10 - .946603 = 9.053397$$

c) The variance of the portfolio return is the variance of the mean return of the stocks in the portfolio. We want this variance to be less than .1 percent. Hence, we have

$$\sigma^2_{\bar{r}} = \frac{\sigma^2}{n} < .1$$

which, when both sides are multiplied by n and divided by .1, yields

$$n > \frac{\sigma^2}{.1} = \frac{4}{.1} = 40.$$

d) We want the difference between the mean and the upper bound to be .1 and, assuming the required sample is large, this distance to be less than $z(.95)$ standard normal deviations. Using XLISPSTAT, we can determine that .95 of the probability mass of the standard normal distribution is to the left of a value of z equal to.

```
> (normal-quant .95)
1.6448536269514717
```

Hence

$$z(.95)\,\frac{\sigma}{\sqrt{n}} = 1.645\,\frac{\sqrt{4}}{\sqrt{n}} = 1.645\,\frac{2}{\sqrt{n}} < .1$$

which, when both sides are multiplied by $\sqrt{n}$ and divided by .1, yields

$$\sqrt{n} > 1.645\,\frac{2}{.1} = 32.90.$$

The required number of stocks is thus 1082.50 which rounds up to 1083. In comparison to the number of stocks on a typical stock exchange, this is a very large number of stocks.

e) The variance of the return to an individual stock in the population is 4 and the standard deviation is therefore 2. The mean return is 10. If we assume that the population is approximately normally distributed, .95 of the probability weight will lie to the left of the upper confidence interval, which will thus equal

$$10 + (1.96)(2) = 10 + 3.92 = 13.92$$

percent. And .05 of the probability mass will lie to the left of the lower confidence interval (.95 will lie above it), which will be

$$10 - (1.96)(2) = 10 - 3.92 = 6.08.$$

[Give yourself 5 points for understanding how to correctly answer each of the five parts of the question.]

Question 5: (10 points)

We want the difference between the true population proportion and the sample population proportion to be equal to .01. Thus, if we center a normal distribution on the sample proportion $\bar{p}$, we want the probability mass between $\bar{p} - .01$ and $\bar{p} + .01$ to be .95. This requires a probability weight of .025 beyond each confidence limit, so the appropriate z-value will be $z(.975) = 1.96$. You can find this z-value by entering into XLISPSTAT the command

```
>(normal-quant .975)
1.9599649845400536
```

Thus, we want

$$1.96 = \frac{.01}{\sigma_{\bar{p}}}$$

where

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}.$$

The maximum possible value of $\sigma_{\bar{p}}$ will occur where $p = .5$. To err on the safe side, we therefore set $p(1-p) = (.5)(.5) = .25$. This yields a value of $\sigma_{\bar{p}}$ equal to $.5/\sqrt{n}$. Hence, we have

$$1.96 = \frac{.01}{.5/\sqrt{n}}.$$

Multiplying both sides by $.5/\sqrt{n}$ and dividing both sides by 1.96 we obtain

$$\frac{.5}{\sqrt{n}} = \frac{.01}{1.96}.$$

Inverting this and multiplying both sides by .5, we obtain

$$\sqrt{n} = \frac{(.5)(1.96)}{.01} = \frac{.98}{.01} = 98.$$

The required sample size is thus $98^2 = 9604$.

[Give yourself 5 points for recognizing how to obtain the appropriate value of z and 5 points for setting up the equation determining $\sqrt{n}$.]

Question 6: (15 points)

The sample of $n = 25$ babies from the population of all babies born at Sunnybrook Hospital yields the statistics $\bar{X} = 6$ and $s = 1.2$. To construct a prediction interval we need to first calculate the variance of a new observation.

$$\begin{aligned}
\sigma^2_{new} &= E\{(X_{new} - \mu)^2\} \\
&= E\{(X_{new} - \bar{X} + \bar{X} - \mu)^2\} \\
&= E\{(X_{new} - \bar{X})^2 + (\bar{X} - \mu)^2\} \\
&= s^2 + \frac{s^2}{n} \\
&= s^2 \left(1 + \frac{1}{n}\right) \\
&= 1.2^2 \left(1 + \frac{1}{25}\right) \\
&= (1.44)(1.04) \ = \ 1.4976
\end{aligned}$$

The standard deviation of a new observation is thus $\sqrt{1.4976} = 1.2238$. With a sample size of 25 we should, to be safe, use the t-distribution in obtaining the confidence interval. Using XLISPSTAT, we obtain $t(.975; 24)$ (remember, we want .025 of the probability mass of the t-distribution on each tail)

```
> (t-quant .975 24)
2.0638985615320604
```

so that

$$U = 6 + (2.06)(1.2238) = 6 + 2.52 = 8.52$$

and

$$L = 6 - (2.06)(1.2238) = 6 - 2.52 = 3.48$$

[Give yourself 5 points for knowing how to approach the problem, 5 points for knowing how to calculate the standard deviation of a new observation, and 5 points for knowing how to correctly calculate calculate the confidence interval.]