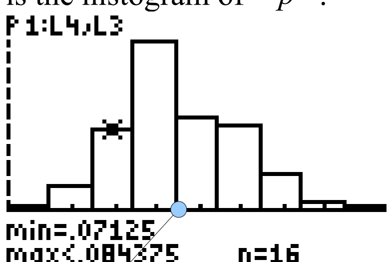
9.8

(a) Here is the frequency table.

X	Freq	P-hat
9	1	0.05
10	0	0.05
11	0	0.06
12	0	0.06
13	3	0.07
14	2	0.07
15	5	0.08
16	11	0.08
17	12	0.09
18	12	0.09
19	9	0.1
20	7	0.1
21	5	0.11
22	6	0.11
23	7	0.12
24	10	0.12
25	4	0.13
26	1	0.13
27	2	0.14
28	2	0.14
29	0	0.15
30	1	0.15

And here is the histogram of $\hat{p}$:



The distribution of the sample proportions is very symmetric.

(c)
$\bar{x}_{\hat{p}}=0.0981$

There is little if any bias in this estimate.

(d) $\mu_{\hat{p}}=0.10$

(e) The mean of the sampling distribution from samples of size 1000 would be the same as the mean of the sampling distribution for samples of size 200, however, the variability of the the new distribution would be notably smaller.

9.10

(a) Graph (a) indicates large bias and large variability.
(b) Graph (b) indicates small bias and small variability.
(c) Graph (c) indicates small bias and large variability.
(d) Graph (d) indicates large bias and small variability.

9.14

Assign digits 0-9: 0-1 represent the occurrence of harmful egg masses; and 2-9 represent no harmful egg masses found.

We use randInt(0, 9, 10) for a simulated SRS of size 10 of square-yard frames. Below are the results of my 20 SRS's, namely the proportion of each that contained harmful insect egg masses.

p-hat				
0.3	0.1	0.3	0.2	0
0.1	0.1	0.2	0.2	0.3
0.4	0	0.2	0.1	0.4
0.4	0	0.1	0.3	0.1

Stem	Leaf					
0	0	0	0			
0	1	1	1	1	1	1
0	2	2	2	2		
0	3	3	3	3		
0	4	4	4			

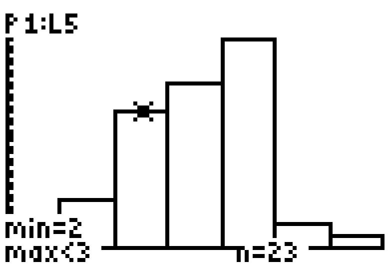
0|1 = 0.1

The sample mean for the above table is 0.19. It is somewhat symmetric.

The mean of the complete sampling distribution is $\mu_p = p = 0.2$

The mean of the sampling distribution of the new field is, of course, $\mu_p = p = 0.4$

9.16



The histogram for the 100 repetitions of samples from $\overline{X}_3$ is shown here.

The distribution is quite symmetric, the mean is $\mu = 3.5$, and the standard deviation is $\sigma_{\overline{X}} = \frac{1.7078}{\sqrt{3}}$.

9.19

(a) Supposing that $p = 0.7$, we know that $\mu_{\hat{p}} = 0.7$ and $\sigma_{\hat{p}} = \sqrt{\frac{(0.7)(0.3)}{1012}} \approx 0.0144$

(b) It is reasonable to assume that 1012 is less than 10% of the population of interest.

(c) Since $\begin{array}{c} 1012(0.7) = 708.4 > 10, \\ 1012(0.3) = 303.6 > 10 \end{array}$ the conditions for using the normal approximation for the distribution of $\hat{p}$ are satisfied

(d) $P(\hat{p} \leq 0.67) =$ normalcdf(-1E99, 0.67, 0.7, 0.0144) = 0.018 = 1.8% This is not a very likely result. We either got a very unusual sample, or our supposition in part (a) that $p = 0.7$ is not correct.

(e) We desire to know n such that,

$$\sqrt{\frac{(0.7)(0.3)}{n}}=\frac{1}{2}\sqrt{\frac{(0.7)(0.3)}{1012}}=\sqrt{\frac{(0.7)(0.3)}{4(1012)}}$$ So $n = 4(1012) = 4048$

(f) I know for sure it would be different than 0.67. My guess is that it would be bigger. I think that more teenagers drink their cereal milk than adults. In general they are hungrier. Also, some adults begin to avoid drinking milk because of the fat content and less of a need for calcium.

9.20

(a and b) We know that $\hat{p}$ is an unbiased estimator of p, so the mean of the distribution of $\hat{p}$ is of course, $p = 0.4$. Since it is reasonable to assume that 1785 < 0.1(The population of U.S. Adults) we can say that the standard deviation of the distribution of $\hat{p}$ is

$$\sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.4)(0.6)}{1785}} = 0.01159 .$$

(c) Since $np = 714 > 10$ and $nq = 1071 > 10$, we can use the normal approximation for the distribution of $\hat{p}$.

(d) $P(0.37 < \hat{p} < 0.43) =$ normalcdf(0.37, 0.43, 0.4, 0.1159) = 0.99

We can see from the calculation above, 99% of the time we will get a sample result for $\hat{p}$ within 3 percentage points of the true proportion of 0.4.

9.21

Sample Size, n	$P(0.37 < \hat{p} < 0.43)$
300	= normalcdf(0.37, 0.43, 0.4, $\sqrt{(0.4)(0.6)/n}$) = 0.7112
1200	= normalcdf(0.37, 0.43, 0.4, $\sqrt{(0.4)(0.6)/n}$) = 0.966
4800	= normalcdf(0.37, 0.43, 0.4, $\sqrt{(0.4)(0.6)/n}$) = 0.999

9.22

(a) Since $np = 70 > 10$ and $nq = 430 > 10$, we can say that the distribution of $\hat{p}$ is approximately normal with mean, $\mu_{\hat{p}} = 0.14$ and $\sigma_{\hat{p}} = \sqrt{(0.14)(0.86)/500}$, assuming 500 < 0.1 (Population of Motorcycle owners).

(b) $P(\hat{p} \geq 0.2) =$ normalcdf(0.2, 1E99, 0.14, $\sqrt{(0.14)(0.86)/500}$) = 0.0000552

(c) $P(\hat{p} \geq 0.2) =$ normalcdf(0.15, 1E99, 0.14, $\sqrt{(0.14)(0.86)/500}$) = 0.2597

9.24

Note that the word accurate here refers to how close we actually get to the actual probability in question. This is an appropriate use of the word accurate.

Since the larger sample size causes the distribution to be "more" normal, the calculation involving the larger sample size will be more accurate. Hence the calculation involving the 500 motorcycle owners will be more accurate than the calculation involving the 100 mail orders.

9.30

(a) This situation violates Rule of Thumb 2: since $np = 15(0.3) = 4.5 < 10$. We cannot use the approximate normal distribution to calculate this probability.

(b) Since 50 > 0.1(316), we violate Rule of Thumb 1. We cannot estimate the standard deviation of the distribution using our $\sqrt{pq/n}$ formula.

(c) $P(X \le 3) =$ binomcdf(15, 0.3, 3) = 0.2968

9.32

Let X represent the scores on the ACT test.

(a) $P(x > 21) =$ normalcdf(21, 1E99, 18.6, 5.9) = 0.342

(b) The mean of the sampling distribution from a sample of size 50 is 18.6, and the standard deviation (assuming that 50 is less than one-tenth the population of test takers) is $5.9/\sqrt{50}$. These values are true whether X is normally distributed or not. What depends on the distribution of X is the *shape* of the sampling distribution.

(c) Since X is normally distributed, we know that the sampling distribution of $\overline{X}$ is too. So

$$P(\bar{x} > 21) = \text{normalcdf}(21,\ 1E99,\ 18.6,\ 5.9/\sqrt{50}) = 0.002$$

9.40

(a) If X represents the lifetime of a set of brake pads, and we are given that X is normally distributed, then $\overline{X}$ is also normally distributed with mean 55,000 miles and standard deviation (assuming that 8 is less than one-tenth the population of all of the new brand of brake pads) is $4500/\sqrt{8}$.

(b) $P(\bar{x} \le 51{,}800) =$ normalcdf(- 1E99, 51800, 55000, $4500/\sqrt{8}$) = 0.022

9.42

Note that the distribution of $\overline{X}$, the sample mean from a sample of 22 children, is approximately normal with mean 13.6 and standard deviation $3.1/\sqrt{22}$.

So $L =$ invnorm(0.05, 13.6, $3.1/\sqrt{22}$) = 12.51

9.44

(b) In my TI-84 simulation, 7 of my 50 groups (14%) that had $\hat{p} \le 0.65$. This is reasonable if p = 0.7. I have no reason to doubt the Customs agents.

(c) $\hat{p} \sim N\left(0.7, \sqrt{\dfrac{(0.7)(0.3)}{100}}\right)$

(d) $P(\hat{p} \le 0.65) =$ normalcdf(-1E99, 0.65, 0.7, .0458) = 0.138 This is very close to my simulated result.

(e) If n = 1000, we have $\hat{p} \sim N\left(0.7, \sqrt{\dfrac{(0.7)(0.3)}{1000}}\right)$, and

$P(\hat{p} \le 0.65) =$ normalcdf(-1E99, 0.65, 0.7, .01449) = 0.00028

9.50

(a) It is more likely that the distribution will be quite skewed to the the right. At rush hour most cars will have just 1 or two persons and a few will have more than that.

(b) Since $n = 700$ which is much bigger than 25 or 30, and since it is reasonable to assume that the population of all cars that will ever pass through this intersection is greater than 7000, we can claim by the CLT that

$$\bar{x} \sim N\left(1.5, \frac{0.75}{\sqrt{700}}\right)$$

(c) Note that 700 cars carrying 1075 people gives $\bar{x} = \frac{1075}{700} = 1.5357$ people per car, so

$P(\bar{x} \geq 1.5357) =$ normalcdf(1.5357, 1E99, 1.5, .0287) = 0.107