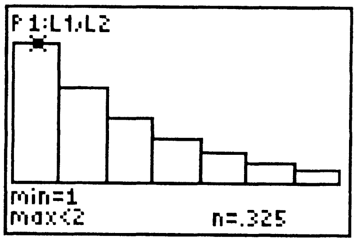
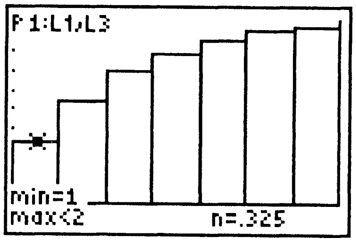




8.64 (a) By the 68–95–99.7 rule, the probability of any one observation falling within the interval $\mu - \sigma$ to $\mu + \sigma$ is .68. Let X = the number of observations out of 5 that fall within this interval. Assuming that the observations are independent, X is $B(5, .68)$. Then, $P(X = 4)$ = `binompdf (5, .68, 4)` = .3421

(b) By the 68–95–99.7 rule, 95% of all observations fall within the interval $\mu - 2\sigma$ to $\mu + 2\sigma$. Thus, 2.5% (half of 5%) of all observations will fall above $\mu + 2\sigma$. Let X = the number of observations that must be taken before we observe one falling above $\mu + 2\sigma$. Then X is geometric with $p = .025$. $P(X = 4) = (1 - .025)^3(.025) = (.975)^3(.025) = .0232$.

8.65 $P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{n}{0} p^0 (1 - p)^{n-0} = 1 - (1)(1)(1 - p)^n = 1 - (1 - p)^n$.

9

Sampling Distributions

9.1 $\mu = 2.5003$ is a parameter; $\bar{x} = 2.5009$ is a statistic.

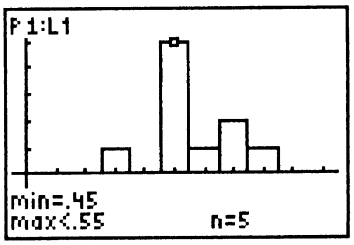
9.2 $\hat{p} = 7.2\%$ is a statistic.

9.3 $\hat{p} = 48\%$ is a statistic; $p = 52\%$ is a parameter.

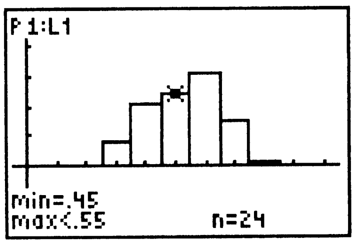
9.4 Both $\bar{x}_1 = 335$ and $\bar{x}_2 = 289$ are statistics.

9.5 (a) Since the proportion of times the toast will land butter-side down is 0.5, the result of 20 coin flips will simulate the outcomes of 20 pieces of falling toast (landing butter-side up or butter-side down).
(b) Answers will vary.
(c) Answers will vary; however, it is more likely that the center of this distribution will be close to 0.5, and it is more likely that the shape will be close to normal.
(d) Answers will vary.
(e) We obtain a more accurate representation of a sampling distribution when many samples are taken.

9.6 (a)

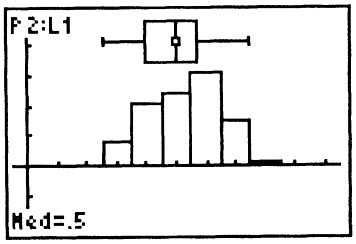


The results appear to be quite variable.
(b)



The center is close to 0.5, and the shape is approximately normal.

(c)

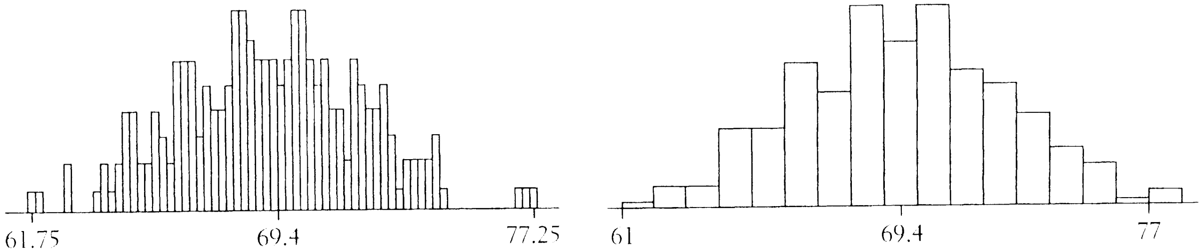


The median and mean are extremely close.

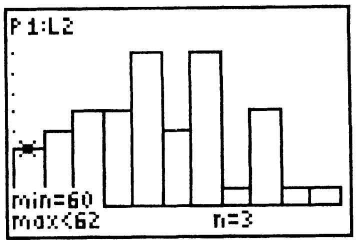
(d) The spread of the distribution did not seem to change. To decrease the spread, I would increase the number of trials, n. For example, use `randBin (50, .5)`.

9.7 (a) The scores will vary depending on the starting row. Note that the smallest possible mean is 61.75 (from the sample 58, 62, 62, 65) and the largest is 77.25 (from 73, 74, 80, 82).

(b)–(c) Answers will vary; shown below are two views of the sampling distribution. The first shows all possible values of the experiment (so the first rectangle is for 61.75, the next is for 62.00, etc.); the other shows values grouped from 61 to 61.75, 62 to 62.75, etc. (which makes the histogram less bumpy). The tallest rectangle in the first picture is 8 units; in the second, the tallest is 28 units.



(d) There are $(10 \times 9)/2 = 45$ possible samples of size 2 that can be drawn from the population.

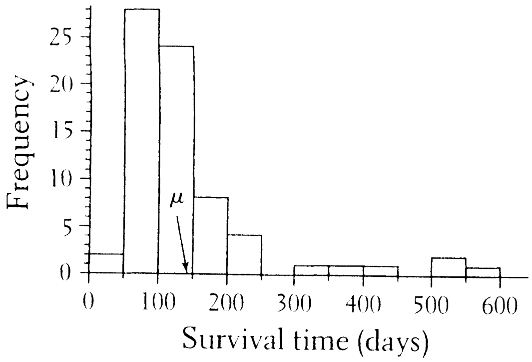


(e) The shapes and centers for the two distributions are roughly the same. However, the spread is a little larger for the distribution corresponding to $n = 2$. This distribution is somewhat more irregular, reflecting the fact that sample means based on samples of size 2 tend to be more variable than those based on samples of size 4.

9.8 (a) Table is on the next page; histogram not shown. (b) The histogram actually does *not* appear to have a normal shape. The sampling distribution is quite normal in appearance, but even a sample of size 100 does not *necessarily* show it. (c) The mean of $\hat{p}$ is 0.0981. The bias seems to be small. (d) The mean of the sampling distribution should be $p = 0.10$. (e) The mean would still be 0.10, but the spread would be smaller.

p	$\hat{p}$	Count	p	$\hat{p}$	Count	p	$\hat{p}$	Count
9	0.045	1	18	0.090	12	24	0.120	10
13	0.065	3	19	0.095	9	25	0.125	4
14	0.070	2	20	0.100	7	26	0.130	1
15	0.075	5	21	0.105	5	27	0.135	2
16	0.080	11	22	0.110	6	28	0.140	2
17	0.085	12	23	0.115	7	30	0.150	1

9.9 (a) Below, left. (b) For the 72 survival times, $\mu = 141.847$ days. (c) Means will vary with samples. (d) It would be unlikely (though not impossible) for all five $\bar{x}$ values to fall on the same side of μ. This is one implication of the unbiasedness of $\bar{x}$: Some values will be higher and some lower than μ. (But note, it is not necessarily half and half.) (e) Shown (below, right) is a stemplot for one set of 100 sample means, which approximates the sampling distribution of $\bar{x}$. This set of means varied from 85 to 225 days and had mean 138.1 and standard deviation 25.9 days. The mean of the (theoretical) sampling distribution would be μ. (f) Answers will vary.



8	5
9	1789
10	1133589999
11	0000111367789
12	245566789
13	01112233666689999
14	00001111223689
15	001223355567889
16	02348
17	234567
18	14
19	223
20	
21	
22	5

9.10 (a) Large bias and large variability. (b) Small bias and small variability. (c) Small bias, large variability. (d) Large bias, small variability.

9.11 (a) Since the smallest number of total tax returns (i.e., the smallest population) is still more than 100 times the sample size, the variability will be (approximately) the same for all states.

(b) Yes, it will change—the sample taken from Wyoming will be about the same size, but the sample in, e.g., California will be considerably larger, and therefore the variability will decrease.

9.12 $\bar{x} = 64.5$ is a statistic; $\mu = 63$ is a parameter.

9.13 $\hat{p} = 4.5\% = .045$ is a statistic.

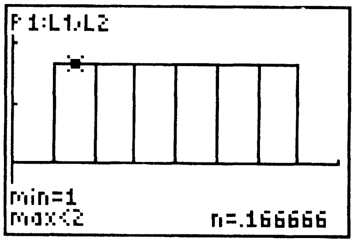
9.14 (a) Use digits 0 and 1 (or any other 2 of the 10 digits) to represent the presence of egg masses. Reading the first 10 digits from line 116, for example, gives YNNNN NNYNN—2 square yards with egg masses, 8 without—so $\hat{p} = 0.2$.

(b) The stemplot *might* look like the one below (which is close to the sampling distribution of $\hat{p}$).

(c) The mean would be $p = 0.2$. (d) 0.4.

0.0	00
0.0	55555
0.1	000000
0.1	5555
0.2	00
0.2	5

9.15 (a) $\mu = 3.5$, $\sigma = 1.708$.

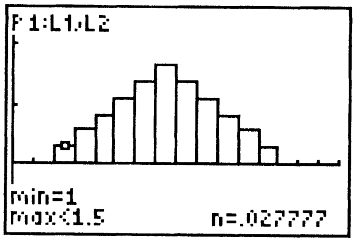


(b) This is equivalent to rolling a pair of fair, six-sided dice.

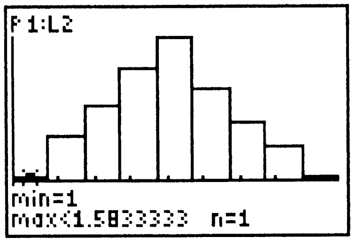
(c)

SRS of size 2	$\bar{x}$
1, 1	1
2, 1 1, 2	1.5
3, 1 2, 2 1, 3	2
4, 1 2, 3 3, 2 1, 4	2.5
5, 1 2, 4 3, 3 4, 2 1, 5	3
6, 1 2, 5 3, 4 4, 3 5, 2 1, 6	3.5
6, 2 3, 5 4, 4 5, 3 2, 6	4
6, 3 4, 5 5, 4 3, 6	4.5
6, 4 5, 5 4, 6	5
6, 5 5, 6	5.5
6, 6	6

(d) Histogram below. The center is identical to that of the population distribution. The shape is normal (symmetric and bell-shaped) rather than uniform. The spread is smaller than that of the population distribution; the probability of observing a value at some distance from the center remains constant for the population distribution but decreases with increasing distance for the histogram corresponding to $n = 2$.



9.16 Answers will vary. A sample histogram is shown below. While the center of the distribution remains the same, the spread is smaller than that of the histogram in Exercise 9.15.



9.17 Assuming that the poll's sample size was less than 780,000–10% of the population of New Jersey—the variability would be practically the same for either population. (The sample size for this poll would have been considerably less than 780,000.)

9.18 (a) The digits 1 to 41 are assigned to adults who say that they have watched *Survivor II*. The program outputs a proportion of "Yes" answers. For (b), (c), (d), and (e), answers will vary; however, as the sample size increases from 5 to 25 to 100, the variability of the sample proportions should decrease.

9.19 (a) $\mu = p = 0.7$; $\sigma = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.7)(0.3)}{1012}} = 0.0144$.

(b) The population (all U.S. adults) is clearly at least 10 times as large as the sample (the 1012 surveyed adults).

(c) $np = (1012)(.7) = 708.4 \geq 10$; $n(1-p) = (1012)(.3) = 303.6 \geq 10$.

(d) $P(\hat{p} \leq .67) = P(Z \leq -0.25) = 0.0186$—this is a fairly unusual result if 70% of the population actually drinks the cereal milk.

(e) Multiply the sample size by 4; we would need to sample (1012)(4) = 4048 adults.

(f) It would probably be higher, since teenagers (and children in general) have a greater tendency to drink the cereal milk.

9.20 (a) $\mu = p = 0.4$, $\sigma = \sqrt{(0.4)(0.6) \div 1785} = 0.0116$. (b) The population (U.S. adults) is considerably larger than 10 times the sample size. (c) $np = 714$, $n(1-p) = 1071$—both are much bigger than 10. (d) $P(0.37 < \hat{p} < 0.43) = P(-2.586 < Z < 2.586) = 0.9904$. Over 99% of all samples should give $\hat{p}$ within ±3% of the true population proportion.

9.21 For $n = 300$: $\sigma = 0.02828$ and $P = 0.7108$. For $n = 1200$: $\sigma = 0.01414$ and $P = 0.9660$. For $n = 4800$: $\sigma = 0.00707$ and $P = 1$ (approximately). Larger sample sizes give more accurate results (the sample proportions are more likely to be close to the true proportion).

9.22 (a) The distribution is approximately normal with mean $\mu = p = 0.14$ and standard deviation $\sigma = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.14)(0.86)}{500}} = 0.0155$. (b) 20% or more Harley owners is unlikely; $P(\hat{p} > 0.20) \approx P(Z > 3.87) < 0.0002$. There is a fairly good chance of finding at least 15% Harley owners; $P(\hat{p} > 0.15) \approx P(Z > 0.64) = 0.2611$.

9.23 (a) 0.86 (86%). (b) We use the normal approximation (Rule of Thumb 2 is *just* satisfied—$n(1-p) = 10$). The standard deviation is 0.03, and $P(\hat{p} \leq 0.86) = P(Z \leq -1.33) = 0.0918$. (*Note:* The exact probability is 0.1239.) (c) Even when the claim is correct, there will be some variation in sample proportions. In particular, in about 10% of samples we can expect to observe 86 or fewer orders shipped on time.

9.24 The calculation for Exercise 9.22 should be more accurate. This calculation is based on a larger sample size (500, as opposed to the 100 of Exercise 9.23). Rule of Thumb 2 is easily satisfied in Exercise 9.22 but just barely satisfied in Exercise 9.23.

9.25 (a) $\mu = p = 0.15$, $\sigma = \sqrt{(0.15)(0.85) \div 1540} = 0.0091$. (b) The population (U.S. adults) is considerably larger than 10 times the sample size (1540). (c) $np = 231$, $n(1 - p) = 1309$—both are much bigger than 10. (d) $P(0.13 < \hat{p} < 0.17) = P(-2.198 < Z < 2.198) = 0.9722$. (e) To achieve $\sigma = .403$, we need a sample *nine* times as large—about 13,860.

9.26 For $n = 200$: $\sigma = 0.02525$, and the probability is $P = 0.5704$. For $n = 800$: $\sigma = 0.01262$ and $P = 0.8858$. For $n = 3200$: $\sigma = 0.0631$ and $P = 0.9984$. Larger sample sizes give more accurate results (the sample proportions are more likely to be close to the true proportion).

9.27 $P(\hat{p} \geq .608) = P\left(Z \geq \frac{.608 - .6}{\sqrt{\frac{(.6)(.4)}{2500}}}\right) = P(Z \geq .8165)$

$= .2071$. (The exact answer is .213.)

9.28 (a) $\mu = 0.52$, $\sigma = 0.02234$. (b) np and $n(1 - p)$ are 260 and 240 respectively. $P(\hat{p} \geq 0.50) = P(Z \geq -0.8951) = 0.8159$.

9.29 (a) $P(\hat{p} \leq 0.70) = P(Z \leq -1.155) = 0.1241$. (b) $P(\hat{p} \leq 0.70) = P(Z \leq -1.826) = 0.0339$. (c) The test must contain 400 questions. (d) The answer is the same for Laura.

9.30 (a) $np = (15)(0.3) = 4.5$—this fails Rule of Thumb 2. (b) The population size (316) is not at least 10 times as large as the sample size (50)—this fails Rule of Thumb 1. (c) $P(X \leq 3) =$ `binomcdf(15, .3, 3)` $= .2969$.

9.31 (a) $\mu = -3.5\%$, $\sigma = 26\%/\sqrt{5} = 11.628\%$. (b) $P(X \geq 5\%) \approx P(Z \geq 0.3269) = .3719$. (c) $P(\bar{x} \geq 5\%) \approx P(Z \geq 0.73) = 0.2327$. (d) $P(\bar{x} < 0) \approx P(Z < 0.30) = 0.6179$. Approximately 62% of all five-stock portfolios lost money.

9.32 (a) $P(X \geq 21) = P(Z \geq 0.4068) = 0.3421$. (b) $\mu = 18.6$, $\sigma = 5.9/\sqrt{50} = 0.8344$. This result is independent of distribution shape. (c) $P(\bar{x} \geq 21) \approx P(Z \geq 2.8764) = 0.0020$.

9.33 (a) $\sigma/\sqrt{3} \doteq 5.7735$ mg. (b) Solve $\sigma/\sqrt{n} = 3$: $\sqrt{n} = \frac{10}{3}$, so $n = 11.1$ or 12. The average of several measurements is more likely than a single measurement to be close to the mean.

9.34 (a) If we choose many samples, the average of the $\bar{x}$-values from these samples will be close to μ. (I.e., $\bar{x}$ is "correct on the average" in many samples.)
(b) The larger sample will give more information, and therefore more precise results; that is, $\bar{x}$ is more likely to be close to the population truth. Also, $\bar{x}$ for a larger sample is less affected by outliers.

9.35 $\bar{x}$ has approximately a $N(1.6, 0.0849)$ distribution; the probability is $P(Z > 4.71)$—essentially 0.

9.36 $\bar{x}$ (the mean return) has approximately a $N(9\%, 4.174\%)$ distribution; $P(\bar{x} > 15\%) = P(Z > 1.437) = 0.9247$; $P(\bar{x} < 5\%) = P(Z < -0.9583) = 0.1690$.

9.37 (a) $N(123, 0.04619)$. (b) $P(Z > 21.65)$—essentially 0.

9.38 (a) Mean: 40.125, standard deviation: 0.001; normality is not needed. (b) No: We cannot complete $p(\bar{x} > 40.127)$ based on a sample of size 4 because the sample size must be *larger* to justify use of the central limit theorem if the distribution type is unknown.

9.39 (a) $P(X < 295) = P(Z < -1) = 0.8413$. (b) $P(\bar{x} < 295) = P(Z < -2.4495) = 0.0072$.

9.40 (a) $N(55000, 4500/\sqrt{8}) = N(55000, 1591)$. (b) $P(Z < -2.011) = 0.0222$.

9.41 (a) $N(2.2, 0.1941)$. (b) $P(\bar{x} < 2) \approx P(Z < -1.0304) = 0.1515$. (c) $P(\bar{x} < \frac{100}{52}) \approx P(Z < -1.4267) = 0.0768$.

9.42 $\mu - 1.645\,\sigma/\sqrt{n} = 12.513$.

9.43 (a) $p = 68\% = .68$ is a parameter; $\hat{p} = 73\% = 0.73$ is a statistic. (b) $\mu = p = 0.68$, $\sigma = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.68(0.32)}{150}} = 0.0381$. (c) $P(\hat{p} \geq .73) \approx P(Z \geq 1.3128) = 0.0946$. There is a 10% (one in ten) chance that an observation of $\hat{p}$ greater than or equal to the observed value of .73 will be seen.

9.44 (a), (b) Answers will vary. In one simulation, we obtained a total of 9 simulated values of $\hat{p}$ less than or equal to 0.65, a percentage of 18%. (c) $\hat{p}$ has an approximately normal distribution. $\mu = p = 0.7$, $\sigma = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.7(0.3)}{100}} = 0.0458$. (d) $P(\hat{p} \leq 0.65) \approx P(Z \leq -1.091) = 0.1376$. This result is reasonably close to the 18% obtained in our simulation. (e) For $n = 1000$, $\hat{p}$ is again approximately normal, with $\mu = p = 0.7$, $\sigma = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.7(0.3)}{1000}} = 0.0145$. $P(\hat{p} \leq 0.65) \approx P(Z \leq -3.45) = 0.0003$. $\hat{p}$ is less variable for larger sample sizes, so the probability of seeing a value of $\hat{p}$ less than or equal to 0.65 decreases.

9.45 $2P(\bar{x} > 27.4) = 2P\left(Z > \frac{27.4 - 25}{7/\sqrt{10}}\right) \approx 2P(Z > 1.084) = 2(0.1392) = .2784$.

9.46 (a) $\hat{p}$ has an approximately normal distribution with $\mu = p = 0.47$, $\sigma = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.47(0.53)}{1025}} = 0.0156$.

(b) The middle 95% of all sample results will fall within $2\sigma \approx 0.0312$ of the mean 0.47, that is, in the interval 0.4388 to 0.5012.

(c) $P(\hat{p} < 0.45) \approx P(Z < -1.283) = 0.0998$.

9.47 The mean loss from fire, by definition, is the long-term average of *many* observations of the random variable X = fire loss. The behavior of X is much less predictable if only a small number of observations are made. If only 12 policies were sold, then the company would have no protection against the large expense that would be incurred if one of the 12 policyholders happened to lose his or her home. If thousands of policies were sold, then the average fire loss for these policies would be far more likely to be close to μ, and the company's profit would not be endangered by the few large fire-loss payments that it would have to make.

9.48 $P(\bar{x} > 260) = P\left(Z > \frac{260 - 250}{300/\sqrt{10000}}\right) \approx P(Z > 3.33) = 0.0004$.

9.49 (a) $P\left(Z > \frac{105 - 100}{15}\right) = P(Z > \frac{1}{3}) = 0.36944$. (b) Mean: 100; standard deviation: 1.93649. (c) $P\left(Z > \frac{105 - 100}{1.93649}\right) = P(Z > 2.5820) = 0.00491$. (d) The answer to (a) could be quite different; (b) would be the same (it does not depend on normality at all). The answer we gave for (c) would still be fairly reliable because of the central limit theorem.

9.50 (a) No—a count assumes only whole-number values, so it cannot be normally distributed. (b) $N(1.5, 0.02835)$. (c) $P\left(\bar{x} > \frac{1075}{700}\right) = P(Z > 1.2599) = 0.10386$.

9.51 $\mu + 2.33\sigma/\sqrt{n} = 1.4625$.

9.52 (a) $np = (25000)(0.141) = 3525$.

(b) $P(X \geq 3500) = P\left(Z \geq \frac{3500 - 3525}{55.027}\right) = P(Z \geq -0.4543) = 0.6752$.

9.53 $P\left(\frac{750}{12} < \bar{x} < \frac{825}{12}\right) = P(-1.732 < Z < 2.598) = 0.95368$.