Chapter 7

7.2 Suppose the mean time devoted to prayer in the United States is to be estimated using data on Catholics, Jews, Muslims, and Protestants. Because these groups are not of equal size, simple random sampling might get none or too few of the smallest group. To avoid this, a number proportional to each group's representation in the population could be selected; thus, variability between the groups (strata) is controlled and precision is increased. This method works best when individuals within the strata are alike but different from those in the other strata. It adds nothing to precision if those in one strata are like those in another.

7.4 For a sample to be a random sample each and every element or set of elements in the population must have an equal chance to be selected into a sample. Those who valued their time at more than $1.00 were less likely to respond.

7.6 Answer depends on the random numbers used by the student.

7.8 If the sample size increases by a multiple of four(from n to 4n, for example), the standard error decreases by a multiple of $\frac{1}{2}$. (Originally $F_{\bar{x}} = F/\sqrt{n}$, but now $F_{\bar{x}} = F/\sqrt{4n} = \frac{1}{2}F/\sqrt{n}$)

7.10 The standard error of the mean is the standard deviation of the mean, which the authors do not acknowledge. It is a measure of the average variability in the sample means around their expected value (which is the population mean). The standard error of the mean is equal to $F/\sqrt{n}$, where F is the standard deviation of the population and n is the sample size. The standard error of the mean is estimated by $s/\sqrt{n}$, where s is the standard deviation of a sample. Gardner and Altman’s definition does not distinguish clearly between F and s; their statement about the standard deviation pertains to s and not F.

7.12 Yes. The standard error of an estimator (like $\bar{x}$) varies inversely with the sample size but increasing the sample size does not result in proportional reductions in the standard error. A k times increase in the sample size results in a $\frac{1}{\sqrt{k}}$ times reduction in the standard error.

7.14 Whenever the sample size is sufficiently large (i.e., $n \geq 30$), the sampling distribution of the mean is approximately normally distributed according to the Central Limit Theorem.

7.16 a. The sampling distribution of the sample mean($\bar{x}$) is normal with a mean of 24 and a standard error of the mean of 4 ($= 4/\sqrt{1} = 4/1$). Thus,

$$P(\bar{x}>28) = P(z>\frac{28-24}{4}) = P(z>1) = 0.1587$$

Or using EXCEL, 0.15866 =1-NORMDIST(28,24,4,1)

b. The sampling distribution of the sample mean($\bar{x}$) is normal with a mean of 24 and a standard error of the mean of 2 ($= 4/\sqrt{4} = 4/2$). Thus,

$$P(\bar{x}>28) = P(z>\frac{28-24}{2}) = P(z>2) = 0.0228$$

Or using EXCEL, 0.02275 =1-NORMDIST(28,24,2,1)

c. The sample size is 1 in a and 4 in b. If n=1, the sampling distribution of the sample mean is just the population distribution itself. With a larger size sample, sampling distribution of the sample mean has smaller standard deviation than that in the population; i.e., as n increases the standard error of the mean decreases.

7.18 Assuming the normality of the population distribution, ensures that the sampling distribution of the sample mean is normal with a mean of 1.5 and a standard error of 0.2. Thus, from EXCEL we have 0.10565 =NORMDIST(1.25,1.5,0.2,1):

$$P(\overline{x} \le 1.25) = P(z \le \frac{1.25 - 1.5}{0.2}) = P(z \le -1.25) = 0.1056$$

Other than assuming that sampling is random we do not need additional assumptions.

7.20 Although the population distribution is not normal, the sampling distribution of the sample mean grade point average is approximately normal(because the sample size 36 is large) with a mean of 2.80 and a standard error of 0.133 = 0.8/√36 Thus, from EXCEL 0.933241 =1-NORMDIST(2.6,2.8,0.1333,1):

$$P(\overline{x} \ge 2.60) = P(z > \frac{2.60 - 2.80}{0.133}) = P(z > -1.5) = 0.9332$$

7.22 a. False. The population proportion is a parameter and the sample proportion is a random variable. There is no reason to believe that any one value of the sample proportions would equal to the population proportion. It is the average(mean) of all the sample proportion values that equals the population proportion.

b. False. The sample size has nothing to do with whether the population proportion is equal to a sample proportion.

c. Yes. See above.

d. False. The variance of the distribution of the sample proportion is B(1-B)/n where B is the population proportion, while the variance of X is F^2.

e. False. s^2/n is an estimator of the variance of the sample mean. It has nothing to do with the variance of the sampling distribution of the sample proportion.

f. Yes. The variance of the sample proportion for samples of size n equals B(1-B)/n where B is the population proportion.

7.24 a. n=40 and B(the probability of "success", i.e., the probability of getting a nurse who is compassionate and caring)=0.9. If the binomial random variable X denotes the number of nurses who are compassionate and caring when 40 nurses are randomly drawn, then the probability of getting 34 or fewer such nurses is

$$P(x \le 34 | n=40, B=0.9) = \sum_{x=0}^{34} \binom{40}{x} 0.9^x (1-0.9)^{40-x} = 0.206$$

Or using EXCEL, 0.20627 =BINOMDIST(34,40,0.9,1)

b. We can think of the distribution of the sample proportion of nurses who are compassionate and caring when we randomly draw 40 nurses. Because n=40 is sufficiently large to invoke the Central Limit Theorem, the sampling distribution of the sample proportion is approximately normal with a mean of 0.9 and a standard error of 0.047 ($= \sqrt{\frac{(0.9)(0.1)}{40}}$). Thus, the probability the sample proportion(p) is 0.85(=34/40) or less is

$$P(p \le 0.85) = P(z \le \frac{0.85-0.9}{0.047}) = P(z \le -1.06) = 0.1446$$

Or using EXCEL, 0.1447 = NORMDIST(0.85,0.9,0.047,1)

7.26 Because division managers handpicked the 35 employees from the pool of 1,700 employees, it cannot be considered to be a random sample. We might suspect that division managers selected those employees who were satisfied and would not want to unionize.

7.28 a. Because n=100 is large enough to invoke the Central Limit Theorem, the sample mean weight will be approximately normally distributed with an expected value of 320 lbs. and a standard error of the mean equal to

5 lbs. ($= 60/\sqrt{144} = 60/12$)

b. The probability that the average weight of 144 people lies beyond 335 lbs. will be

$$P(\bar{x} > 335) = P(z > \frac{335-320}{5}) = P(z > 3) = 0.0013$$

c. If n=36, the standard error of the mean would be 10 lbs. ($= 60/\sqrt{36} = 60/6$)

7.30 Since the sample size n=16 is not large, we need to assume that the population distribution is normal to answer the question. Under this assumption, the sample mean width will be normally distributed with a mean of 0.25 inches and a standard error of the mean equal to 0.0000025 inches

$(= 0.00001/\sqrt{16} = 0.00001/4)$. Thus, the probability that a sample mean is 0.25007 or larger will be

$$P(\bar{x} \geq 0.25007) = P(z \geq \frac{0.25007-0.25}{0.0000025}) = P(z \geq 28) = 0.0000$$

7.32 The population is normally distributed with a mean of 11.7 and a standard error of 0.40. Thus, the sampling distribution of the sample mean is normal with a mean of 11.7 and a standard error of 0.1(= $0.40/\sqrt{16}=0.40/4$). We are asked to find the value of b such that P($\bar{x}$<b)= 0.025. Using EXCEL, 11.504 =NORMINV(0.025,11.7,0.1) so the unreported value of the sample mean is 11.504 ounces.

7.34 Even if the population mean(μ) is negative, say negative 1 percent, we might observe a sample mean($\bar{x}$) of 5.5 percent or greater with the following probability:

$$P(\bar{x} \geq 0.055) = P(z \geq \frac{0.055-(-0.01)}{0.045}) = P(z \geq 1.44) = 0.0749$$

This probability is not small. Thus, it could be possible to observe 5.5 percent or greater as a monthly change in sales of new homes when the sales in fact decreased by 1 percent.

7.36 It is known to us that 0.35 ≤ P(30<x<36) ≤ 0.40, where X, denoting men's waist size, is a normal random variable with 33 as its mean. Normality of X suggests that the following should be satisfied:

$P(30<x<36) = P(\frac{30-33}{F}<z<\frac{36-33}{F}) = P(-\frac{3}{F}<z<\frac{3}{F})$ is between 0.35 and 0.40 where F is the standard deviation of the distribution of men's waist size.

This implies $0.455 < \frac{3}{F} < 0.525$; the implied value of F should lie between 5.7 and 6.6. Thus, the limits on the implied standard error of the mean for a sample of size 4 are $2.85 < \frac{F}{\sqrt{4}} < 3.3$

7.38 Because the distribution of the time to be "up and running" is highly left skewed as shown in the diagram, it might be appropriate to use the binomial distribution instead of the normal distribution even though the sample size is relatively large.

We can compare the following probabilities:

If the binomial random variable X denotes the number of "successes" with B=0.98 when 30 days are randomly drawn, then the probability of getting 29 or fewer "successes" is

$$P(x \le 29 | n = 30, B = 0.98) = \sum_{x=0}^{29} \binom{30}{x} 0.98^{x}(1-0.98)^{30-x} = 0.455$$

using EXCEL, 0.454516 =BINOMDIST(29,30,0.98,1)

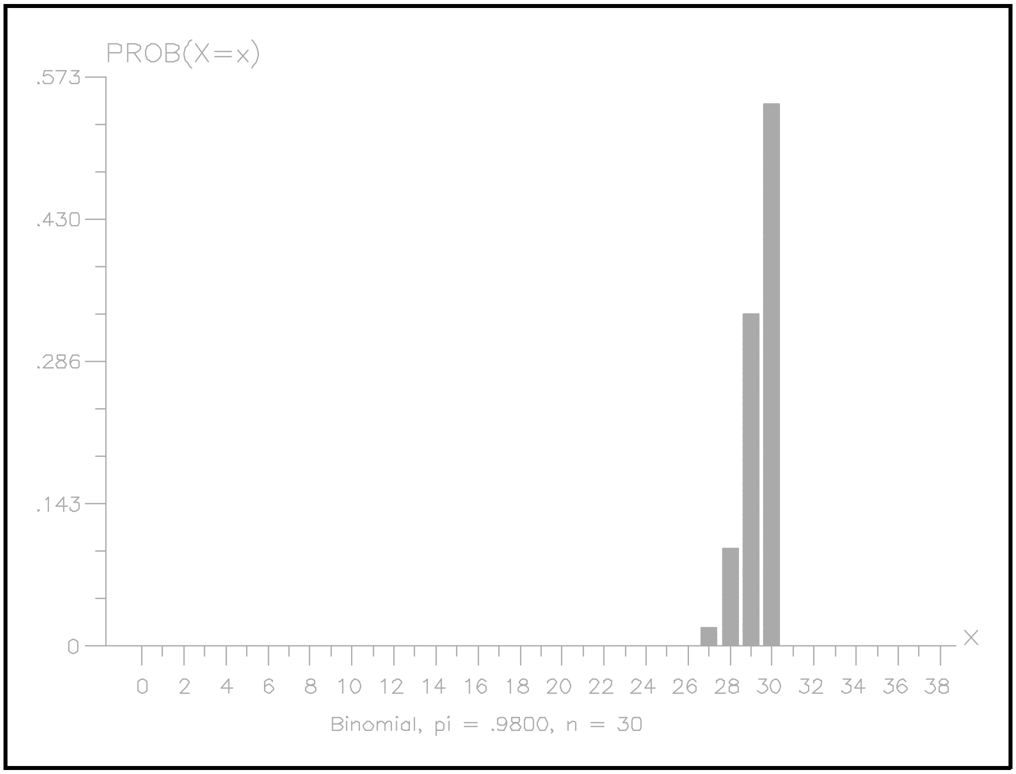
Or as an alternative method, we could try a calculation with the normal distribution. The sampling distribution of the sample proportion(p) is approximately normal with a mean of

0.98 and a standard deviation of 0.0256 ($= \sqrt{\frac{(0.98)(0.02)}{30}}$)

Thus, the probability that the sample proportion(p) is 0.9667(=29/30) or less will be

$$P(p \le 0.9667) = P(z \le \frac{0.9667 - 0.98}{0.0256}) = P(z \le -0.52) = 0.3015$$

These two probabilities do not agree because n=30 is not sufficiently large for a proportion so far from B = 0.5. The probability 0.4545 calculated using the binomial distribution is the more appropriate in this case.



7.40 The mean difference is calculated to be 7.83 miles as follows:

SKI AREA	ACTUAL MILES OF TRAILS	MILES CLAIMED BY RESORT	CLAIMED LESS ACTUAL
Killington	70	77	7
Sugarbush	46	53	7
Sunday River	29	36	7
Sugarloaf	36	45	9
Okemo	25	32	7
Smuggler's Notch	22	32	10
COLUMN SUM	228	275	47
MEAN	38	45.83	7.83

$P(X \geq 7.83) \approx 0$ =1-NORMDIST(7.83,0,2/SQRT(6),1)

7.42 a. Under the assumption of normality of the distribution of the average line speed(X), the following probability calculation can be done:

$$P(x \leq 11250) = P(z \leq \frac{11250-11500}{142}) = P(z \leq -1.76) = 0.0392$$

= NORMDIST(11250,11500,142,1)

Because the line speed is a continuous random variable that can be expected to vary symmetrically around its mean, with values closer to the mean occurring more often, it seems reasonable to assume that the average line speed(X) is normally distributed.

b. Because n=36 is large enough to invoke the Central Limit Theorem, the sample average line speed will be approximately normally distributed with a mean of 11,500 cans per minute and a standard deviation of 23.67 (= $142/\sqrt{36}$) cans per minute. Thus,

$$P(\bar{\bar{x}} \leq 11250) = P\left(z \leq \frac{11250-11500}{142/\sqrt{36}}\right) = P(z \leq -10.56) = 0.0000$$

= NORMDIST(11250,11500,23.67, 1)