Chapter 8

8.2 The words "confidence" and "probability" are interchangeable before sampling, when a sample statistic has not yet been calculated. Once a confidence interval has been constructed, however, the words are no longer interchangeable because the value of the random variable has been realized and thus there is no random variable to which probability applies.

8.4 a. No, there is no reason to believe that two demographers will have the same point estimate for μ. Because they take their own samples independently, they are likely to have different samples and different sample means.

b. It depends on the situation. If F is known to them, then the margin of error $(= \pm z_{0.025}\frac{F}{\sqrt{n}} = \pm 1.96\frac{F}{\sqrt{64}})$ will be the same for both demographers.

However, if F is unknown and thus they use their own sample standard deviations(s) instead of F, then the margin of error $(= \pm t_{(0.025, df=63)}\frac{s}{\sqrt{n}} = \pm 1.998\frac{s}{\sqrt{64}})$ will not be the same because their sample standard deviations may differ. [Note: 1,998 = TINV(0.05,63)]

8.6 The upper end = $\bar{x} + z_{"/2}\frac{F}{\sqrt{n}} = \bar{x} + z_{"/2}\frac{12}{\sqrt{16}} = \bar{x} + 3z_{"/2}$

The lower end = $\bar{x} - z_{"/2}\frac{F}{\sqrt{n}} = \bar{x} - z_{"/2}\frac{12}{\sqrt{16}} = \bar{x} - 3z_{"/2}$

The width (or length) from the lower to the upper end is

$(\bar{x} + 3z_{"/2}) - (\bar{x} - 3z_{"/2}) = 6z_{"/2}$ which was found to be 18.60.

Thus, $z_{"/2}$ = 3.10(=18.6/6). This implies that "/2 = 0.001 and thus " = 0.002. The corresponding level of confidence is 99.8 percent.

8.8 For a confidence interval for the population proportion B, the margin of error is

$$z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = z_{\alpha/2}\sqrt{\frac{(0.73)(0.27)}{767}} = 0.016 z_{\alpha/2}$$

which was calculated to be 0.0413. This implies the value of z is $z_{\alpha/2}$ = 2.58

8.10 The margin of error $\pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$ depends on the value of the sample proportion p. Thus, the margin of error for each question would not be the same as long as the sample proportion for each question differs.

8.12 These cannot be done easily with Appendix Table A.4, but using EXCEL we get

a. P(|t| > 2.1|df=21)= 0.047988479 =TDIST(2.1,21,2)
b. P(t>-2.6|df=15)= 0.9899505 =1-TDIST(2.6,15,1)
c. P(t<b|df=17) = 0.025, b = 2.109818524 =-TINV(0.05,17)
d. P(t>b|df=25) = 0.05, b= 1.708140189 =TINV(0.1,25)

8.14 Because the t distribution has more variability, the confidence interval using the t distribution has a wider interval than the interval using the z distribution, other things being equal(i.e., for a given level of confidence and a given sample size).

8.16 If the distribution of $\bar{x}$ is normal and the population standard deviation(F) is known, the z distribution has to be used. But even in such a situation, the t distribution can also be used instead of z, if the sample size is large enough. The two distributions are pretty much the same for a large sample size.

$$P(\bar{x} \geq 22340.833) = P(z \geq \frac{22340.833-21750}{849.618}) = P(z \geq 0.70) = 0.242$$

8.18 Using the ASCII data file EX8-18.PRN provided on the computer disk, the mean and the standard deviation of the given sample data can be calculated $\bar{x}$=22,340.833, s=2,081.131 respectively. Under the assumption of normality of the population of the car prices, the distribution of $\bar{x}$ is normal with a mean of 21,750 and a standard deviation of 849.618 $(= F/\sqrt{n} = 2081.131/\sqrt{6})$ assuming F ≃ s. Thus, and from EXCEL we have 0.242 =1-NORMSDIST(0.7)

However, there may be no reason to assume that F ≈ s, in which case the t distribution must be used. Using EXCEL we have the following in that case

	A	
1	19500.00	
2	25550.00	
3	21995.00	
4	22500.00	
5	23500.00	
6	21000.00	
7	22340.83	=AVERAGE(A1:A6)
8	2081.13	=STDEV(A1:A6)
9	0.70	=(A7-21750)/(A8/SQRT(6))
10	0.2564	=TDIST(A9,6,1)

Thus, by the t distribution $P(\bar{x} \geq 22340.83) = 0.26$, which is only slightly higher than the 0.24 obtained using the z distribution.

8.20 a. The point estimate of mean profit, i.e., the value of the sample mean is 6.28 cents per dollar of sales.

b. The estimated standard error of mean profit is 0.2794

$(= s/\sqrt{n} = 1.34/\sqrt{23})$

c. We need to assume that the population is normally distributed so that the distribution of $\bar{x}$ can be normal. We have to use the t distribution in the construction of a confidence interval because the population standard deviation is unknown and the sample size n=23 is not large. Thus, a 95 percent confidence interval can be constructed as follows:

$$\bar{x} - t_{(0.025,df=22)}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{(0.025,df=22)}\frac{s}{\sqrt{n}}$$

$$6.28 - 2.074(0.2794) < \mu < 6.28 + 2.074(0.2794)$$

$$5.70 < \mu < 6.86$$

8.22 A 95 percent confidence interval for the population proportion of Indiana school bus drivers with blemished driving records will be 22.86% to 41.14% of all bus drivers.

$$p - z_{0.025}\sqrt{\frac{p(1-p)}{n}} < B < p + z_{0.025}\sqrt{\frac{p(1-p)}{n}}$$

$$0.32 - 1.96\sqrt{\frac{(0.32)(0.68)}{100}} < B < 0.32 + 1.96\sqrt{\frac{(0.32)(0.68)}{100}}$$

$$0.32 - 0.0914 < B < 0.32 + 0.0914$$

$$0.2286 < B < 0.4114$$

8.24 A 95 percent confidence interval for the population proportion of bosses who are significant helpers to women executives is

$$p - z_{0.025}\sqrt{\frac{p(1-p)}{n}} < B < p + z_{0.025}\sqrt{\frac{p(1-p)}{n}}$$

$$0.86 - 1.96\sqrt{\frac{(0.86)(0.14)}{120}} < B < 0.86 + 1.96\sqrt{\frac{(0.86)(0.14)}{120}}$$

$$0.86 - 0.062 < B < 0.86 + 0.062$$

$$0.798 < B < 0.922$$

8.26 A 95 percent confidence interval for the population proportion with faulty brakes is 51% to 69% of all trucks.

$$p - z_{0.025}\sqrt{\frac{p(1-p)}{n}} < B < p + z_{0.025}\sqrt{\frac{p(1-p)}{n}}$$

$$0.61 - 1.96\sqrt{\frac{(0.61)(0.39)}{114}} < B < 0.61 + 1.96\sqrt{\frac{(0.61)(0.39)}{114}}$$

$$0.61 - 0.09 < B < 0.61 + 0.09$$

$$0.52 < B < 0.70$$

8.28 a. According to the sample size determination formula, the implied original sample size n=100 must satisfy the following:

$$100 = [\frac{z_{0.025}F}{M}]^2 = [\frac{1.96F}{M}]^2$$

Thus, $\frac{F}{M} = 5.102$

Now the new sample size with $z_{0.005}$ will be

$$[\frac{z_{0.005}F}{M}]^2 = [\frac{2.575F}{M}]^2 = [2.575(5.102)]^2 = 172.6$$

Thus, n=173 is required.

b. If the population standard deviation is twice as large as what was originally assumed, then the correct sample size should be

$$n = [\frac{z_{0.025}2F}{M}]^2 = [\frac{(1.96)2F}{M}]^2 = [2(1.96)(5.102)]^2 = 400$$

That is, she needs a sample size four times as large as what was originally required.

c. For a margin of error of ± 125, the implied original sample size n must satisfy the following:

$$100 = [\frac{z_{0.025}F}{125}]^2 = [\frac{1.96F}{125}]^2 \quad \text{Thus, } F = 637.7$$

Now the new sample size with a margin of error of 200 will be

$$[\frac{z_{0.025}F}{200}]^2 = [\frac{1.96(637.7)}{200}]^2 = 39.06$$

Thus, n=40 is required.

8.30 From the given information, our best guess of p prior to sampling is 0.85 and the margin of error is ± 0.05. Thus,

$$n = [\frac{z_{0.05}}{M}]^2 p^*(1-p^*) = [\frac{1.645}{0.05}]^2(0.85)(0.15) = 138.01$$

We must consider a random sample of size 139 or more.

8.32 a. Neither is appropriate because the distribution of sample mean is not normal in this situation.

b. z is appropriate. Because the population distribution is normal with known standard deviation, the sampling distribution of the sample mean is also normal with known standard deviation.

c. Neither is appropriate because the distribution of sample mean is not normal in this situation.

d. t is appropriate because the population standard deviation is unknown although the sampling distribution of the sample mean is normal.

e. The t is more appropriate but either may be used if the sample size is large.

f. z is appropriate because the sample size is large (thus the central limit theorem gives approximate normality to $\bar{x}$) and the population standard deviation is known.

8.34 The pay distribution is usually highly right skewed with a few high paying jobs pulling up the mean.

8.36 This confidence interval of $29,700 ± $410 does not say that chances are 95 percent that population average salary is within ± $410 of the survey result($29,700). We can say only that under repeated sampling, 95 percent of similarly constructed intervals would contain the true population average and the interval 29,700 ± 410 may (or may not) be one of those intervals.

8.38 The minimum and maximum SAT scores are 400 and 1600. An approximate value of the population standard deviation F can be calculated to be

$F^* = \frac{1600-400}{5} = 240$. Thus, the required sample size can be

calculated as follows:

$$n = [\frac{z_{0.025}F^*}{M}]^2 = [\frac{1.96(240)}{15}]^2 = 983.4$$

A sample size 984 or more is required to be 95 percent confident that the estimate of the mean improvement is within 15 points of the true mean.

8.40 a. Using the ASCII data file EX8-40.PRN provided on the computer disk, the sample mean($\bar{x}$) is calculated to be 54.48 days.

b. Because the population standard deviation F is unknown, we need to use the t distribution for the interval construction. Using the sample standard deviation s = 24.019 which can be calculated from the data file, the 90 percent confidence interval will be

$$\bar{x} - t_{(0.05, df=24)}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{(0.05, df=24)}\frac{s}{\sqrt{n}}$$

$$54.48 - 1.711\frac{24.019}{\sqrt{25}} < \mu < 54.48 + 1.711\frac{24.019}{\sqrt{25}}$$

$$54.48 - 8.22 < \mu < 54.48 + 8.22$$

$$46.26 < \mu < 62.70$$

Alternatively, using the descriptive statistics routine in EXCEL:

	A	B	C
1	42	Column1	
2	59		
3	30	Mean	54.48
4	35	Standard Error	4.803859559
5	40	Median	55
6	39	Mode	55
7	101	Standard Deviation	24.0192978
8	55	Sample Variance	576.9266667
9	92	Kurtosis	-0.862598362
10	55	Skewness	0.14878756
11	81	Range	89
12	37	Minimum	12
13	73	Maximum	101
14	87	Sum	1362
15	12	Count	25
16	55	Confidence Level(90.0%)	8.218838368
17	63		
18	23	Lower limit	46.26116163 =C3-C16
19	72	Upper limit	62.69883837 =C3+C16
20	66		
21	71		
22	41		
23	20		
24	31		
25	82		

c. There is a truncation or censoring problem in the calculation of the mean and the confidence interval if there are bills that are never paid. If we could remove this effect of truncation, the correct mean time would be greatly increased.

8.42 Using the ASCII data file EX8-42.PRN provided on the computer disk, the sample mean and the sample standard deviation are calculated to be $1180.245 and $1795.443 respectively. Thus, a 95 percent confidence interval will be

$$\bar{x} - t_{(0.025,df=19)}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{(0.025,df=19)}\frac{s}{\sqrt{n}}$$

$$1180.245 - 2.093\frac{1795.443}{\sqrt{20}} < \mu < 1180.245 + 2.093\frac{1795.443}{\sqrt{20}}$$

$$1180.245 - 840.284 < \mu < 1180.245 + 840.284$$

$$339.961 < \mu < 2020.529$$

8.44 The minimum and maximum price of the fir are \$164 and \$206 per 1,000 board feet. An approximate value of the population standard deviation F can be calculated to be

$F^{*} = \frac{206-164}{5} = 8.4$ using the rule of thumb for guessing F

value. Thus, the required sample size can be calculated as follows:

$$n = [\frac{z_{0.025}F^{*}}{M}]^2 = [\frac{1.96(8.4)}{10}]^2 = 2.7$$

We must consider a random sample of size 3 or more.

8.46 A 95 percent confidence interval for the population proportion of such CEOs will be

$$p - z_{0.025}\sqrt{\frac{p(1-p)}{n}} < B < p + z_{0.025}\sqrt{\frac{p(1-p)}{n}}$$

$$0.65 - 1.96\sqrt{\frac{(0.65)(0.35)}{515}} < B < 0.65 + 1.96\sqrt{\frac{(0.65)(0.35)}{515}}$$

$$0.65 - 0.04 < B < 0.65 + 0.04$$

$$0.61 < B < 0.69$$

8.48 a. A 95 percent confidence interval for the population proportion of such customers will be

$$p - z_{0.025}\sqrt{\frac{p(1-p)}{n}} < \text{B} < p + z_{0.025}\sqrt{\frac{p(1-p)}{n}}$$

$$0.83 - 1.96\sqrt{\frac{(0.83)(0.17)}{12}} < \text{B} < 0.83 + 1.96\sqrt{\frac{(0.83)(0.17)}{12}}$$

$$0.83 - 0.21 < \text{B} < 0.83 + 0.21$$

$$0.62 < \text{B} < 1.04$$

b. The interval constructed in part a is suspect because a proportion cannot exceed 1. The sample size 12 is not sufficiently large for the distribution of the sample proportion(p) to be treated as approximately normal.

8.50 The "unbiasedness" as used in this quote pertains to an appraiser not having a personal interest in a property. Statistical unbiasedness pertains to the mean of an estimator equaling the population parameter estimated.

8.52 The point estimate of the population proportion is the sample proportion, which is 0.07, and the point estimate of the standard deviation of the distribution of the sample proportion is 0.012($= \sqrt{\frac{(0.07)(0.93)}{434}}$)

8.54 a. The mean and the standard deviation of the given sample data can be calculated $\hat{x}$ = $521.2, s = $44.561 respectively. Thus, if the true population mean is $517.40, the probability of getting this sample mean or larger can be calculated to be

$$P(\bar{x} \geq 521.2) = P\left(t_{(df=4)} \geq \frac{521.2-517.40}{44.561/\sqrt{5}}\right)$$

$$= P(t_{(df=4)} \geq 0.191) = 0.429$$

b. In the probability calculation in part a, we need to assume that the population distribution is normal so that the sampling distribution of sample mean can be normal.

c. At a typical confidence level 95 percent, a confidence interval for the population mean receipt can be constructed as follows:

$$\bar{\bar{x}} - t_{(0.025, df=4)} \frac{s}{\sqrt{n}} < \mu < \bar{\bar{x}} + t_{(0.025, df=4)} \frac{s}{\sqrt{n}}$$

$$521.2 - 2.776 \frac{44.561}{\sqrt{5}} < \mu < 521.2 + 2.776 \frac{44.561}{\sqrt{5}}$$

$$521.2 - 55.32 < \mu < 521.2 + 55.32$$

$$465.88 < \mu < 576.52$$

This interval includes the Healthy's claimed value $517.40. This suggests the claimed value could be consistent with the true average receipts of the company and thus the IRS does not need to pursue the company.

8.56 The margin of error is calculated to be ±5.88(=11.76/2).

Thus, $z_{"/2} \frac{F}{\sqrt{n}} = z_{"/2} \frac{15}{\sqrt{25}} = 5.88$

This implies $z_{"/2} = 1.96$; the level of confidence is 95 percent.

8.58 A 90 percent confidence interval for the true population proportion of such female judges will be

$$p - z_{0.05} \sqrt{\frac{p(1-p)}{n}} < B < p + z_{0.05} \sqrt{\frac{p(1-p)}{n}}$$

$$0.54 - 1.645 \sqrt{\frac{(0.54)(0.46)}{105}} < B < 0.54 + 1.645 \sqrt{\frac{(0.54)(0.46)}{105}}$$

$$0.54 - 0.08 < B < 0.54 + 0.08$$

$$0.46 < B < 0.62$$

This interval includes not only the majority(more than 50 percent) of judges but also 50 percent or less. This implies that the majority of judges does not necessarily think sexual bias is widespread in California state courts.

8.60 As stated in the solution for Exercise 7.38, because the distribution of the time to be "up and running" is highly left skewed, it would not be appropriate to use the normal distribution in a confidence interval construction even though the sample size(30) is relatively large.

8.62 The margin of error is 5 percentage points in either direction of the poll result. Using p as the sample proportion(poll result), a 95 percent(as implied by the statement "chances are 19 to 20") classical confidence interval for B is p ± 5%. This confidence interval of p ± 5%, however, does not say that chances are 19 to 20 (95 percent) that population proportion B(proportion of black-owned business with annual revenue of at least $100,000) is within 5 percent of the poll result. We can say again that in classical statistics the population proportion is a parameter so there is no probability or chance associated with this interval estimate.

8.64 The 95 percent confidence interval for the mean profit per partner is $0.77 million to $1.04 million, with a margin of error equal to $0.134 million and point estimate of $0.9 million, as shown in the following EXCEL printout:

A	B	C	D
1	1.400	Column1	
2	1.225		
3	1.200	Mean	0.903666667
4	1.185	Standard Error	0.062528977
5	0.940	Median	0.82
6	0.930	Mode	#N/A
7	0.885	Standard Deviation	0.242173688
8	0.820	Sample Variance	0.058648095
9	0.805	Kurtosis	-0.499158536
10	0.800	Skewness	0.777476617
11	0.740	Range	0.775
12	0.700	Minimum	0.625
13	0.655	Maximum	1.4
14	0.645	Sum	13.555
15	0.625	Count	15
16		Confidence Level(95.0%)	0.134111438
17			
18		Lower limit	0.769555229=C3-C16
19		Upper limit	1.037778104=C3+C16
20		Margin of Error	0.134111438=C16

8.66 a. As described in Exercise 8.65, the margin of error for large surveys does not depend on the percentage giving a specific answer on any one question. It is determined by $1/\sqrt{n}$, which in this case is $1/\sqrt{400} = 0.05$.