§ 8.3 Types of Estimators

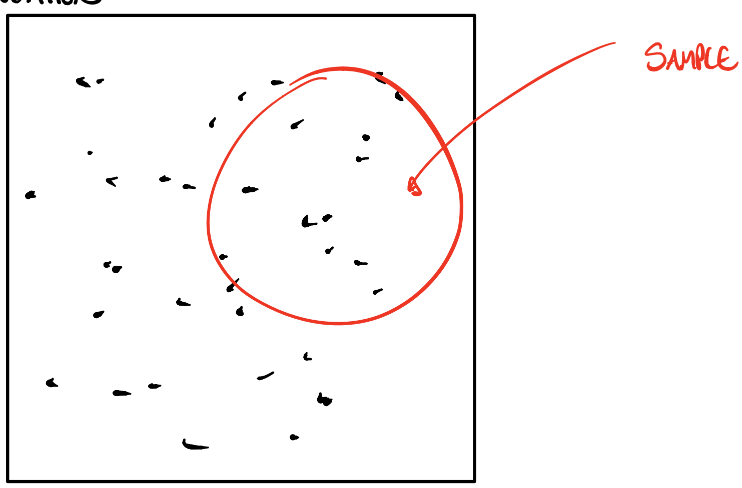
Population Parameters		Sample Statistics	
Mean	μ	Mean	$\bar{x}$
Standard Deviation	σ	Standard Deviation	s
Proportion	p	Proportion	$\hat{p}$

Good at estimating

Unbiased

Use sample statistics to estimate population parameters.

Population



Inferential Statistics

Definition An **estimator** is a rule, usually expressed as a formula, that tells us how to calculate an estimate based on information in the sample.

Estimators are used in two different ways:

- **Point estimation:** Based on sample data, a single number is calculated to estimate the population parameter. The rule or formula that describes this calculation is called the **point estimator,** and the resulting number is called a **point estimate.**

SINGLE #

- **Interval estimation:** Based on sample data, two numbers are calculated to form an interval within which the parameter is expected to lie. The rule or formula that describes this calculation is called the **interval estimator,** and the resulting pair of numbers is called an **interval estimate** or **confidence interval.**

INTERVAL

EXAMPLE 8.3

A veterinarian wants to estimate the average weight gain per month of 4-month-old golden retriever pups that have been placed on a lamb and rice diet. The *population* consists of the weight gains per month of all 4-month-old golden retriever pups that are given this particular diet. The veterinarian wants to estimate the unknown parameter μ, the average monthly weight gain for this *hypothetical* population.

One possible *estimator* based on sample data is the sample mean, $\bar{x} = \Sigma x_i/n$. It could be used in the form of a single number or *point estimate*—for instance, 3.8 pounds—or you could use an *interval estimate* and estimate that the average weight gain will be between 2.7 and 4.9 pounds.

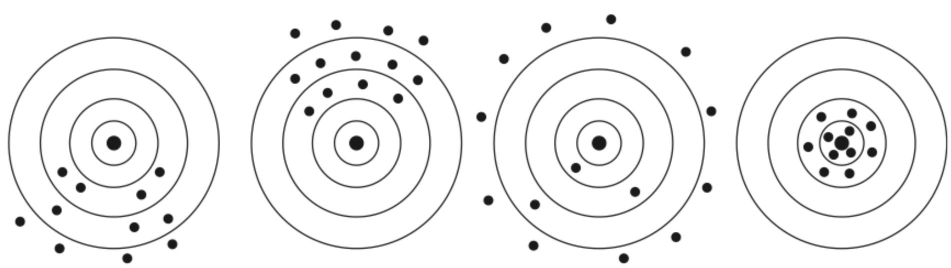
POINT ESTIMATE

INTERVAL ESTIMATE.

§8.4 POINT ESTIMATION

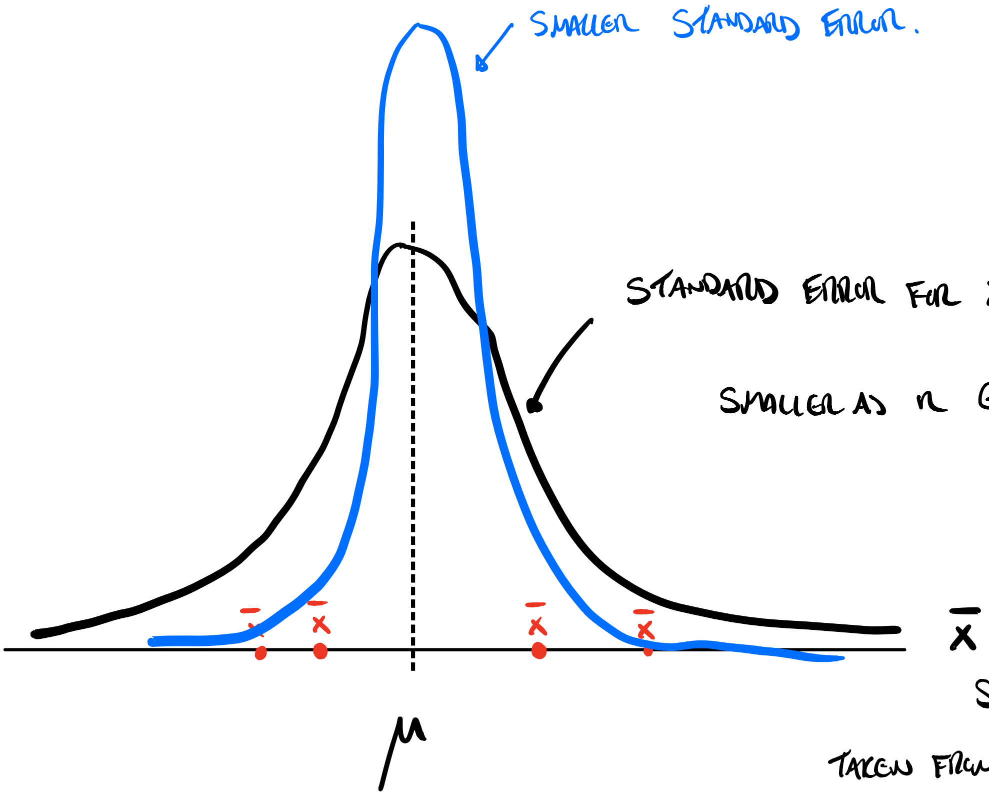
Sampling distributions provide information that can be used to select the **best estimator.** What characteristics would be valuable? First, the **sampling distribution of the point estimator should be centered over the true value of the parameter to be estimated.** That is, the estimator should not constantly underestimate or overestimate the parameter of interest. Such an estimator is said to be **unbiased.**

Definition An estimator of a parameter is said to be **unbiased** if the mean of its distribution is equal to the true value of the parameter. Otherwise, the estimator is said to be **biased.**



Mostly below bull's-eye	Mostly above bull's-eye	Off bull's-eye by a wide margin	Best marksmanship

CLT.



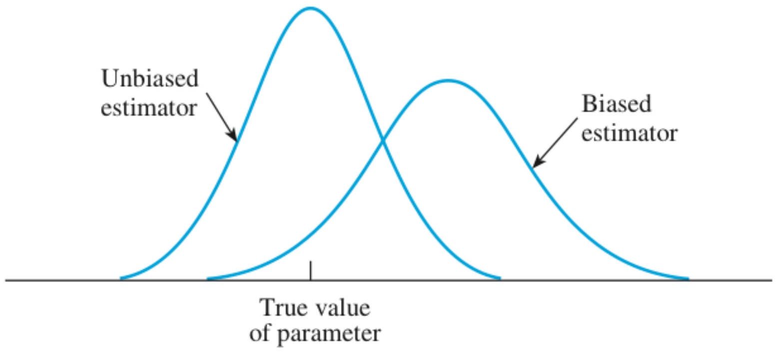
STANDARD ERROR FOR X $= \dfrac{\sigma}{\sqrt{n}}$

SMALLER AS n GETS LARGER

SAMPLE MEANS

TAKEN FROM POPULATION WITH

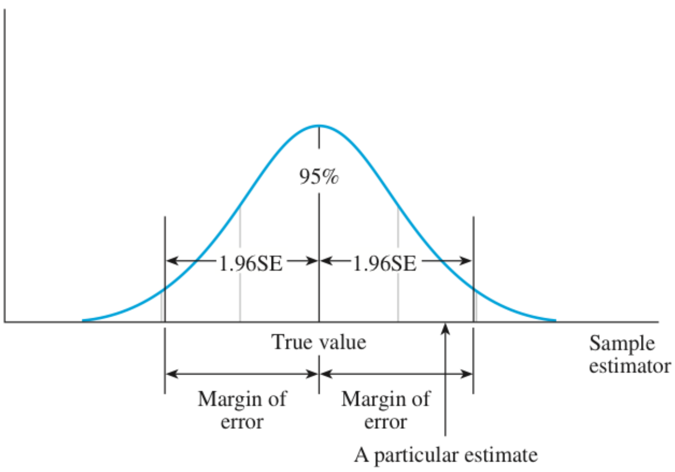
MEAN μ & S.D. σ.



Definition The distance between an estimate and the estimated parameter is called the **error of estimation.**

In this chapter, you may assume that the sample sizes are always large and, therefore, that the *unbiased* estimators you will study have sampling distributions that can be approximated by a normal distribution (because of the Central Limit Theorem). Remember that, for any point estimator with a normal distribution, the Empirical Rule states that approximately 95% of all the point estimates will lie within two (or more exactly, 1.96) standard deviations of the mean of that distribution.

For *unbiased* estimators, this implies that the difference between the point estimator and the true value of the parameter will be less than 1.96 standard deviations or 1.96 standard errors (SE). This quantity, called the 95% **margin of error** (or simply the **"margin of error"**), provides a practical upper bound for the error of estimation (see Figure 8.4). It is possible that the error of estimation will exceed this margin of error, but that is very unlikely.



POINT ESTIMATION OF A POPULATION PARAMETER

- Point estimator: a statistic calculated using sample measurements
- 95% Margin of error: 1.96 × Standard error of the estimator

PREVIOUSLY (CH. 7 CLT.)

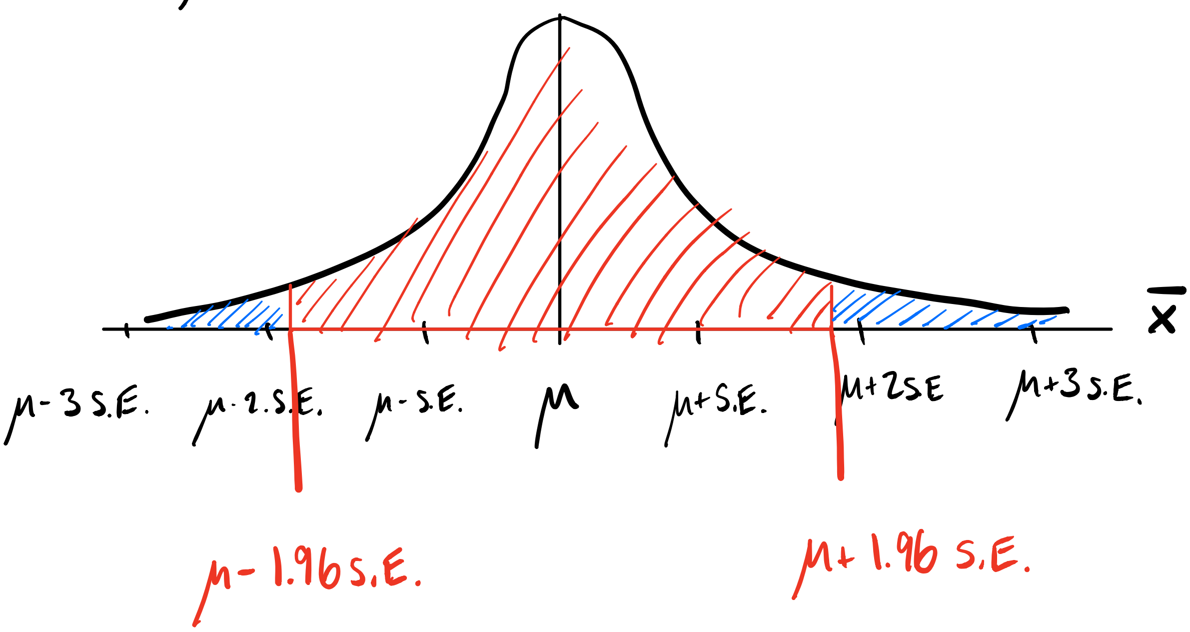
GIVEN A POPULATION WITH MEAN μ & STANDARD DEV. σ.

WHEN SAMPLES OF SIZE $n \geq 30$ ARE SELECTED & SAMPLE MEANS $\bar{x}$ ARE RECORDED.

95% OF SAMPLE MEANS LIE BETWEEN

$$\mu - 1.96 \text{ S.E.} \quad \& \quad \mu + 1.96 \text{ S.E.}$$

$$\mu - 1.96 \frac{\sigma}{\sqrt{n}} \quad \& \quad \mu + 1.96 \frac{\sigma}{\sqrt{n}}.$$



i.e. $P(-1.96 \leq z \leq 1.96) = .95$

i.e. $P(z \leq -1.96) = .025$

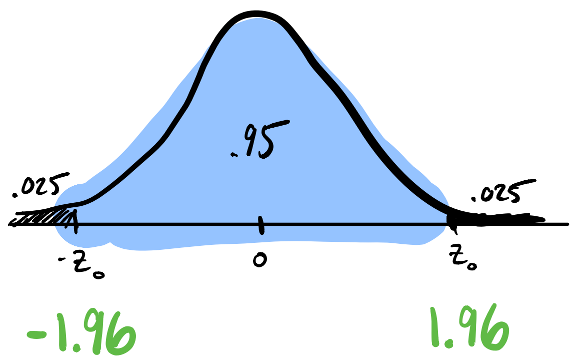


TABLE 3 Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559

THE STANDARD DEVIATION FOR SAMPLE MEANS $\bar{X}$ IS CALLED THE STANDARD ERROR.

Suppose a sample of n=200 bacteria is collected and the time it takes for each bacteria to split in two is recorded. Suppose the mean value for these recorded times is 44 min, and the standard deviation for these recorded times is s=4 min.

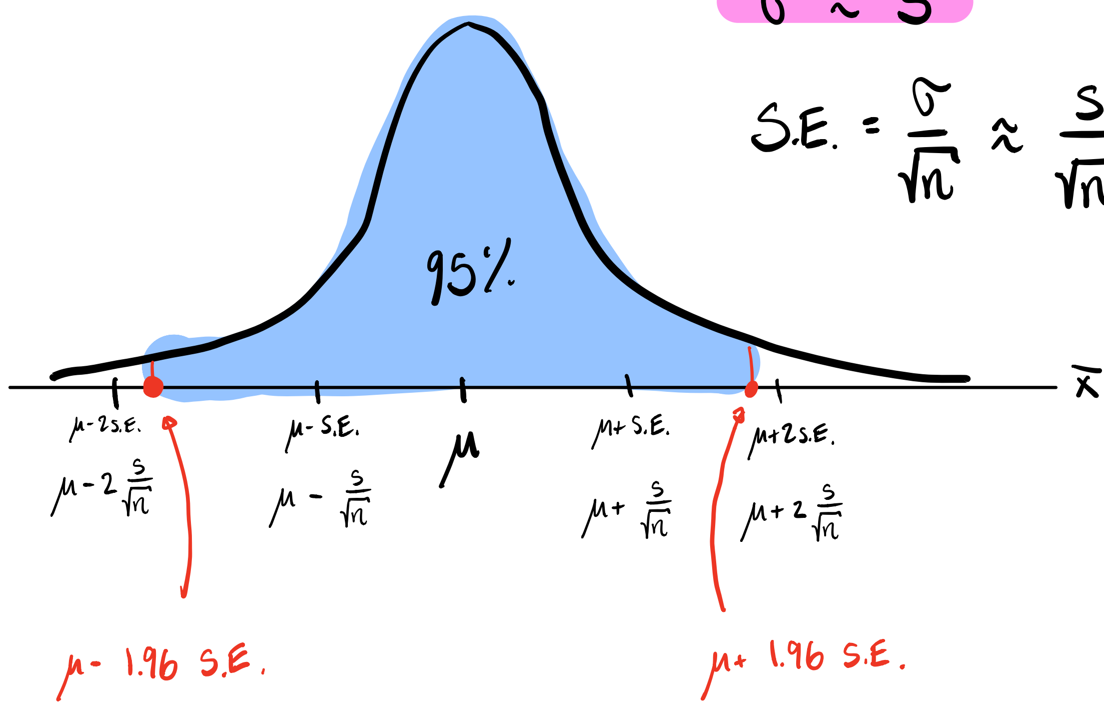
* Central limit theorem says that sample means for sufficiently large samples are approximately normally distributed, with mean equal to the population mean. μ ← UNKNOWN

If we estimate the population standard deviation for the population of ALL bacteria to be the standard deviation of our sample, 4 min. Then...

$\sigma \approx s$ POINT ESTIMATE!

*

$$S.E. = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$$



C.L.T. SAYS PROBABILITY THAT SAMPLE MEAN $\bar{X}$ IS WITHIN 1.96 S.E. OF POPULATION MEAN μ IS 95%.

SAME THING.

95% OF ALL SAMPLE MEANS LIE WITHIN 1.96 S.E. OF POPULATION MEAN.

THE PROBABILITY THAT POPULATION MEAN μ IS WITHIN 1.96 S.E. OF SAMPLE MEAN $\bar{X}$ IS 95%.

$$P(\bar{x} - 1.96\ S.E. \le \mu \le \bar{X} + 1.96\ S.E.) = .95$$

$$44 - 1.96 \frac{4}{\sqrt{200}} \qquad\qquad 44 + 1.96 \frac{4}{\sqrt{200}}$$

$$\underbrace{43.45\ \text{MW} \qquad\qquad 44.55\ \text{MW}}$$

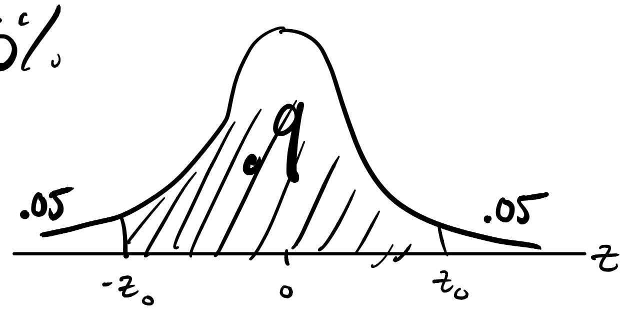
INTERVAL ESTIMATE.

95% SURE THAT POP. MEAN μ IS IN THIS INTERVAL.

$$P(43.45 \le \mu \le 44.55) = .95$$

MIN, MAX → 750, 850

CONFIDENCE → 75%



POP. MEAN

CONFIDENCE LEVEL	CONFIDENCE INTERVAL FOR μ
90%	$[\bar{x} - 1.645\ S.E.\ ,\ \bar{x} + 1.645\ S.E.]$
95%	$[\bar{x} - 1.96\ S.E.\ ,\ \bar{x} + 1.96\ S.E.]$
98%	$[\bar{x} - 2.33\ S.E.\ ,\ \bar{x} + 2.33\ S.E.]$
99%	$[\bar{x} - 2.58\ S.E.\ ,\ \bar{x} + 2.58\ S.E.]$

$$S.E. = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$$

TABLE 3 Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505 *	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559

$* \quad z \quad \left.\begin{array}{l} -1.64 \\ -1.65 \end{array}\right\} -1.645$

NEED TO KNOW...

How to Estimate a Population Mean or Proportion

- To estimate the population mean μ for a quantitative population, the point estimator $\bar{x}$ is *unbiased* with standard error estimated as

$$\text{SE} = \frac{s}{\sqrt{n}}^{\dagger}$$

The 95% margin of error when $n \geq 30$ is estimated as

$$\pm 1.96\left(\frac{s}{\sqrt{n}}\right)$$

- To estimate the population proportion p for a binomial population, the point estimator $\hat{p} = x/n$ is *unbiased,* with standard error estimated as

$$\text{SE} = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

The 95% margin of error is estimated as

$$\pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Assumptions: $n\hat{p} > 5$ and $n\hat{q} > 5$.

EXAMPLE 8.4

An environmentalist is conducting a study of the polar bear, a species found in and around the Arctic Ocean. Their range is limited by the availability of sea ice, which they use as a platform to hunt seals, the mainstay of their diet. The destruction of its habitat on the Arctic ice, which has been attributed to global warming, threatens the bear's survival as a species; it may become extinct within the century.[1] A random sample of $n = 50$ polar bears produced an average weight of 980 pounds with a standard deviation of 105 pounds. Use this information to estimate the average weight of all Arctic polar bears.

EXAMPLE 8.5

In addition to the average weight of the Arctic polar bear, the environmentalist from Example 8.4 is also interested in the opinions of adults on the subject of global warming. In particular, he wants to estimate the proportion of adults who think that global warming is a very serious problem. In a random sample of $n = 100$ adults, 73% of the sample indicated that global warming is a very serious problem. Estimate the true population proportion of adults who believe that global warming is a very serious problem, and find the margin of error for the estimate.

TABLE 8.1 Some Calculated Values of $\sqrt{pq}$

p	pq	$\sqrt{pq}$	p	pq	$\sqrt{pq}$
.1	.09	.30	.6	.24	.49
.2	.16	.40	.7	.21	.46
.3	.21	.46	.8	.16	.40
.4	.24	.49	.9	.09	.30
.5	.25	.50			

Table 8.1 shows how the numerator of the standard error of $\hat{p}$ changes for various values of p. Notice that, for most values of p—especially when p is between .3 and .7—there is very little change in $\sqrt{pq}$, the numerator of SE, reaching its maximum value when $p = .5$. This means that the margin of error using the estimator $\hat{p}$ will also be a maximum when $p = .5$. Some pollsters routinely use the maximum margin of error—often called the **sampling error**—when estimating p, in which case they calculate

$$1.96 \text{ SE} = 1.96\sqrt{\frac{.5(.5)}{n}} \quad \text{or sometimes} \quad 2 \text{ SE} = 2\sqrt{\frac{.5(.5)}{n}}$$

Gallup, Harris, and Roper polls generally use sample sizes of approximately 1000, so their margin of error is

$$1.96\sqrt{\frac{.5(.5)}{1000}} = .031 \qquad \text{or approximately } 3\%$$

In this case, the estimate is said to be within ± 3 percentage points of the true population proportion.