though, seems to contradict that assumption: Samples used in opinion surveys are invariably drawn *without replacement*, in which case X is hypergeometric, not binomial. The consequences of that particular "error," however, are easily corrected and frequently negligible.

It can be shown mathematically that the expected value of $\frac{X}{n}$ is the same regardless of whether X is binomial or hypergeometric; its variance, though, is different. If X is binomial,

$$\text{Var}\left(\frac{X}{n}\right) = \frac{p(1-p)}{n}$$

If X is hypergeometric,

$$\text{Var}\left(\frac{X}{n}\right) = \frac{p(1-p)}{n}\left(\frac{N-n}{N-1}\right)$$

where N is the total number of subjects in the population.

Since $\frac{N-n}{N-1} < 1$, the actual variance of $\frac{X}{n}$ is somewhat smaller than the (binomial) variance we have been assuming, $\frac{p(1-p)}{n}$. The ratio $\frac{N-n}{N-1}$ is called the *finite correction factor.* If N is much larger than n, which is typically the case, then the magnitude of $\frac{N-n}{N-1}$ will be so close to 1 that the variance of $\frac{X}{n}$ is equal to $\frac{p(1-p)}{n}$ for all practical purposes. Thus the "binomial" assumption in those situations is more than adequate. Only when the sample is a sizeable fraction of the population do we need to include the finite correction factor in any calculations that involve the variance of $\frac{X}{n}$. ■

Questions

5.3.1. A commonly used IQ test is scaled to have a mean of 100 and a standard deviation of $\sigma = 15$. A school counselor was curious about the average IQ of the students in her school and took a random sample of fifty students' IQ scores. The average of these was $\bar{y} = 107.9$. Find a 95% confidence interval for the student IQ in the school.

5.3.2. The production of a nationally marketed detergent results in certain workers receiving prolonged exposures to a *Bacillus subtilis* enzyme. Nineteen workers were tested to determine the effects of those exposures, if any, on various respiratory functions. One such function, air-flow rate, is measured by computing the ratio of a person's forced expiratory volume (FEV_1) to his or her vital capacity (VC). (Vital capacity is the maximum volume of air a person can exhale after taking as deep a breath as possible; FEV_1 is the maximum volume of air a person can exhale in one second.) In persons with no lung dysfunction, the "norm" for FEV_1/VC ratios is 0.80. Based on the following data (164), is it believable that exposure to the *Bacillus subtilis* enzyme has no effect on the FEV_1/VC ratio? Answer the question by constructing a 95% confidence interval. Assume that FEV_1/VC ratios are normally distributed with $\sigma = 0.09$.

Subject	FEV_1/VC	Subject	FEV_1/VC
RH	0.61	WS	0.78
RB	0.70	RV	0.84
MB	0.63	EN	0.83
DM	0.76	WD	0.82
WB	0.67	FR	0.74
RB	0.72	PD	0.85
BF	0.64	EB	0.73
JT	0.82	PC	0.85
PS	0.88	RW	0.87
RB	0.82		

5.3.3. Mercury pollution is widely recognized as a serious ecological problem. Much of the mercury released into the environment originates as a byproduct of coal burning and other industrial processes. It does not become dangerous until it falls into large bodies of water, where microorganisms convert it to methylmercury (CH_3^{203}), an organic form that is particularly toxic. Fish are the intermediaries: They ingest and absorb the methylmercury and are then eaten by humans. Men and women, however, may not metabolize CH_3^{203} at the same rate. In one study investigating that issue, six women were given a known amount of protein-bound methylmercury. Shown in the following table are the half-lives of the methylmercury in their systems (114). For men, the average CH_3^{203} half-life is believed to be eighty days. Assume that for both genders, CH_3^{203} half-lives are normally distributed with a standard deviation (σ) of eight days. Construct a 95% confidence interval for the true female CH_3^{203} half-life. Based on these data, is it believable that males and females metabolize methylmercury at the same rate? Explain.

Females	CH_3^{203} Half-Life
AE	52
EH	69
LJ	73
AN	88
KR	87
LU	56

5.3.4. A physician who has a group of thirty-eight female patients aged 18 to 24 on a special diet wishes to estimate the effect of the diet on total serum cholesterol. For this group, their average serum cholesterol is 188.4 (measured in mg/100mL). Because of a large-scale government study, the physician is willing to assume that the total serum cholesterol measurements are normally distributed with standard deviation of $\sigma = 40.7$. Find a 95% confidence interval of the mean serum cholesterol of patients on the special diet. Does the diet seem to have any effect on their serum cholesterol, given that the national average for women aged 18 to 24 is 192.0?

5.3.5. Suppose a sample of size n is to be drawn from a normal distribution where σ is known to be 14.3. How large does n have to be to guarantee that the length of the 95% confidence interval for μ will be less than 3.06?

5.3.6. What "confidence" would be associated with each of the following intervals? Assume that the random variable Y is normally distributed and that σ is known.

(a) $\left(\bar{y} - 1.64 \cdot \frac{\sigma}{\sqrt{n}}, \bar{y} + 2.33 \cdot \frac{\sigma}{\sqrt{n}}\right)$

(b) $\left(-\infty, \bar{y} + 2.58 \cdot \frac{\sigma}{\sqrt{n}}\right)$

(c) $\left(\bar{y} - 1.64 \cdot \frac{\sigma}{\sqrt{n}}, \bar{y}\right)$

5.3.7. Five independent samples, each of size n, are to be drawn from a normal distribution where σ is known. For each sample, the interval $\left(\bar{y} - 0.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{y} + 1.06 \cdot \frac{\sigma}{\sqrt{n}}\right)$ will be constructed. What is the probability that at least four of the intervals will contain the unknown μ?

5.3.8. Suppose that $y_1, y_2, \ldots, y_n$ is a random sample of size n from a normal distribution where σ is known. Depending on how the tail-area probabilities are split up, an infinite number of random intervals having a 95% probability of containing μ can be constructed. What is unique about the particular interval $\left(\bar{y} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{y} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right)$?

5.3.9. If the standard deviation (σ) associated with the pdf that produced the following sample is 3.6, would it be correct to claim that

$$\left(2.61 - 1.96 \cdot \frac{3.6}{\sqrt{20}}, 2.61 + 1.96 \cdot \frac{3.6}{\sqrt{20}}\right) = (1.03, 4.19)$$

is a 95% confidence interval for μ? Explain.

2.5	0.1	0.2	1.3
3.2	0.1	0.1	1.4
0.5	0.2	0.4	11.2
0.4	7.4	1.8	2.1
0.3	8.6	0.3	10.1

5.3.10. In 1927, the year he hit sixty home runs, Babe Ruth batted .356, having collected 192 hits in 540 official at-bats (140). Based on his performance that season, construct a 95% confidence interval for Ruth's probability of getting a hit in a future at-bat.

5.3.11. To buy a thirty-second commercial break during the telecast of Super Bowl XXIX cost approximately $1,000,000. Not surprisingly, potential sponsors wanted to know how many people might be watching. In a survey of 1015 potential viewers, 281 said they expected to see less than a quarter of the advertisements aired during the game. Define the relevant parameter and estimate it using a 90% confidence interval.

5.3.12. During one of the first "beer wars" in the early 1980s, a taste test between Schlitz and Budweiser was the focus of a nationally broadcast TV commercial. One hundred people agreed to drink from two unmarked mugs and indicate which of the two beers they liked better; fifty-four said, "Bud." Construct and interpret the corresponding 95% confidence interval for p, the true proportion of beer drinkers who prefered Budweiser to Schlitz. How would Budweiser and Schlitz executives each have put these results in the best possible light for their respective companies?

5.3.13. The Pew Research Center did a survey of 2253 adults and discovered that 63% of them had broadband Internet connections in their homes. The survey report noted that this figure represented a "significant jump" from the similar figure of 54% from two years earlier. One way to define "significant jump" is to show that the earlier number does not lie in the 95% confidence interval. Was the increase significant by this definition?

Source: http://www.pewinternet.org/Reports/2009/10-Home-Broadband-Adoption-2009.aspx.

5.3.14. If (0.57, 0.63) is a 50% confidence interval for p, what does $\frac{k}{n}$ equal and how many observations were taken?

5.3.15. Suppose a coin is to be tossed n times for the purpose of estimating p, where $p = P(\text{heads})$. How large must n be to guarantee that the length of the 99% confidence interval for p will be less than 0.02?

5.3.16. On the morning of November 9, 1994—the day after the electoral landslide that had returned Republicans to power in both branches of Congress—several key races were still in doubt. The most prominent was the Washington contest involving Democrat Tom Foley, the reigning speaker of the house. An Associated Press story showed how narrow the margin had become (120):

> With 99 percent of precincts reporting, Foley trailed Republican challenger George Nethercutt by just 2,174 votes, or 50.6 percent to 49.4 percent. About 14,000 absentee ballots remained uncounted, making the race too close to call.

Let $p = P(\text{Absentee voter prefers Foley})$. How small could p have been and still have given Foley a 20% chance of overcoming Nethercutt's lead and winning the election?

5.3.17. Which of the following two intervals has the greater probability of containing the binomial parameter p?

$$\left[\frac{X}{n} - 0.67\sqrt{\frac{(X/n)(1-X/n)}{n}}, \frac{X}{n} + 0.67\sqrt{\frac{(X/n)(1-X/n)}{n}}\right]$$

$$\text{or} \quad \left(\frac{X}{n}, \infty\right)$$

5.3.18. Examine the first two derivatives of the function $g(p) = p(1-p)$ to verify the claim on p. 305 that $p(1-p) \leq \frac{1}{4}$ for $0 < p < 1$.

5.3.19. The financial crisis of 2008 highlighted the issue of excessive compensation for business CEOs. In a Gallup poll in the summer of 2009, 998 adults were asked, "Do you favor or oppose the federal government taking steps to limit the pay of executives at major companies?", with 59% responding in favor. The report of the poll noted a margin of error of ± 3 percentage points. Verify the margin of error and construct a 95% confidence interval.

Source: http://www.gallup.com/poll/120872/Americans-Favor-Gov-Action-Limit-Executive-Pay.aspx.

5.3.20. Viral infections contracted early during a woman's pregnancy can be very harmful to the fetus. One study found a total of 86 deaths and birth defects among 202 pregnancies complicated by a first-trimester German measles infection (45). Is it believable that the true proportion of abnormal births under similar circumstances could be as high as 50%? Answer the question by calculating the margin of error for the sample proportion, 86/202.

5.3.21. Rewrite Definition 5.3.1 to cover the case where a finite correction factor needs to be included (i.e., situations where the sample size n is not negligible relative to the population size N).

5.3.22. A public health official is planning for the supply of influenza vaccine needed for the upcoming flu season. She took a poll of 350 local citizens and found that only 126 said they would be vaccinated.

- **(a)** Find the 90% confidence interval for the true proportion of people who plan to get the vaccine.
- **(b)** Find the confidence interval, including the finite correction factor, assuming the town's population is 3000.

5.3.23. Given that n observations will produce a binomial parameter estimator, $\frac{X}{n}$, having a margin of error equal to 0.06, how many observations are required for the proportion to have a margin of error half that size?

5.3.24. Given that a political poll shows that 52% of the sample favors Candidate A, whereas 48% would vote for Candidate B, and given that the margin of error associated with the survey is 0.05, does it make sense to claim that the two candidates are tied? Explain.

5.3.25. Assume that the binomial parameter p is to be estimated with the function $\frac{X}{n}$, where X is the number of successes in n independent trials. Which demands the larger sample size: requiring that $\frac{X}{n}$ have a 96% probability of being within 0.05 of p, or requiring that $\frac{X}{n}$ have a 92% probability of being within 0.04 of p?

5.3.26. Suppose that p is to be estimated by $\frac{X}{n}$ and we are willing to assume that the true p will not be greater than 0.4. What is the smallest n for which $\frac{X}{n}$ will have a 99% probability of being within 0.05 of p?

5.3.27. Let p denote the true proportion of college students who support the movement to colorize classic films. Let the random variable X denote the number of students (out of n) who prefer colorized versions to black and white. What is the smallest sample size for which the probability is 80% that the difference between $\frac{X}{n}$ and p is less than 0.02?

5.3.28. University officials are planning to audit 1586 new appointments to estimate the proportion p who have been incorrectly processed by the payroll department.

- **(a)** How large does the sample size need to be in order for $\frac{X}{n}$, the sample proportion, to have an 85% chance of lying within 0.03 of p?
- **(b)** Past audits suggest that p will not be larger than 0.10. Using that information, recalculate the sample size asked for in part (a).