6 Confidence Intervals

- Using Statistics
- Confidence Interval for the Population Mean When the Population Standard Deviation is Known
- Confidence Intervals for μ When σ is Unknown - The *t* Distribution
- Large-Sample Confidence Intervals for the Population Proportion
- The Finite-Population Correction Factor
- Confidence Intervals for the Population Variance
- Sample Size Determination
- One-Sided Confidence Intervals
- Using the Computer
- Summary and Review of Terms

© The McGraw-Hill Companies, Inc., 1999

6-1 Introduction

- **Consider the following statements:**
  - **$\bar{x}$ = 550**
    - A single-valued estimate that conveys little information about the actual value of the population mean.
  - **We are 99% confident that μ is in the interval [449,551]**
    - An interval estimate which locates the population mean within a narrow interval, with a high level of confidence.
  - **We are 90% confident that μ is in the interval [400,700]**
    - An interval estimate which locates the population mean within a broader interval, with a lower level of confidence.

© The McGraw-Hill Companies, Inc., 1999

Types of Estimators

- Point Estimate
  - A single-valued estimate.
  - A single element chosen from a sampling distribution.
  - Conveys little information about the actual value of the population parameter, about the accuracy of the estimate.
- Confidence Interval or Interval Estimate
  - An interval or range of values believed to include the unknown population parameter.
  - Associated with the interval is a measure of the ***confidence*** we have that the interval does indeed contain the parameter of interest.

© The McGraw-Hill Companies, Inc., 1999

Confidence Interval or Interval Estimate

A ***confidence interval*** or ***interval estimate*** is a range or interval of numbers believed to include an unknown population parameter. Associated with the interval is a measure of the ***confidence*** we have that the interval does indeed contain the parameter of interest.

- A confidence interval or interval estimate has two components:
  - A range or interval of values
  - An associated *level of confidence*

© The McGraw-Hill Companies, Inc., 1999

6-2 Confidence Interval for μ When σ Is Known

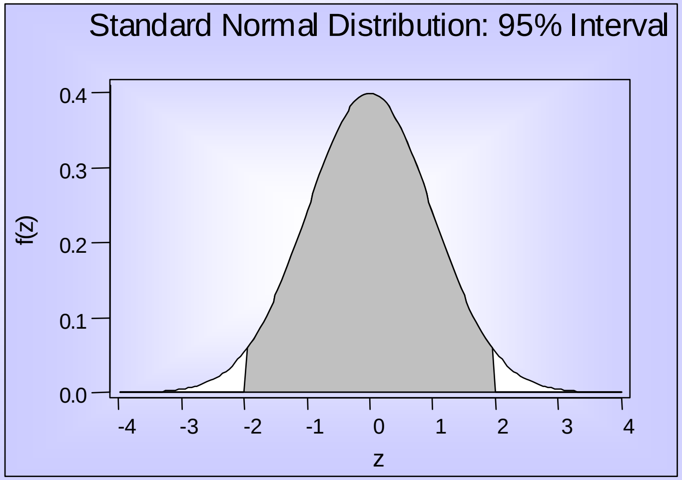
- If the population distribution is normal, *the sampling distribution of the mean is normal*.
- If the sample is sufficiently large, regardless of the shape of the population distribution, *the sampling distribution is normal* (Central Limit Theorem).

In either case:

$$P\left(\mu - 1.96\frac{\sigma}{\sqrt{n}} < \bar{x} < \mu + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

or

$$P\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$



© The McGraw-Hill Companies, Inc., 1999

6-2 Confidence Interval for μ when σ is Known (Continued)

Before sampling, there is a 0.95probability that the interval

$$\mu \pm 1.96\frac{\sigma}{\sqrt{n}}$$

will include the sample mean (and 5% of the it won't).

Conversely, after sampling, approximately 95% of such intervals

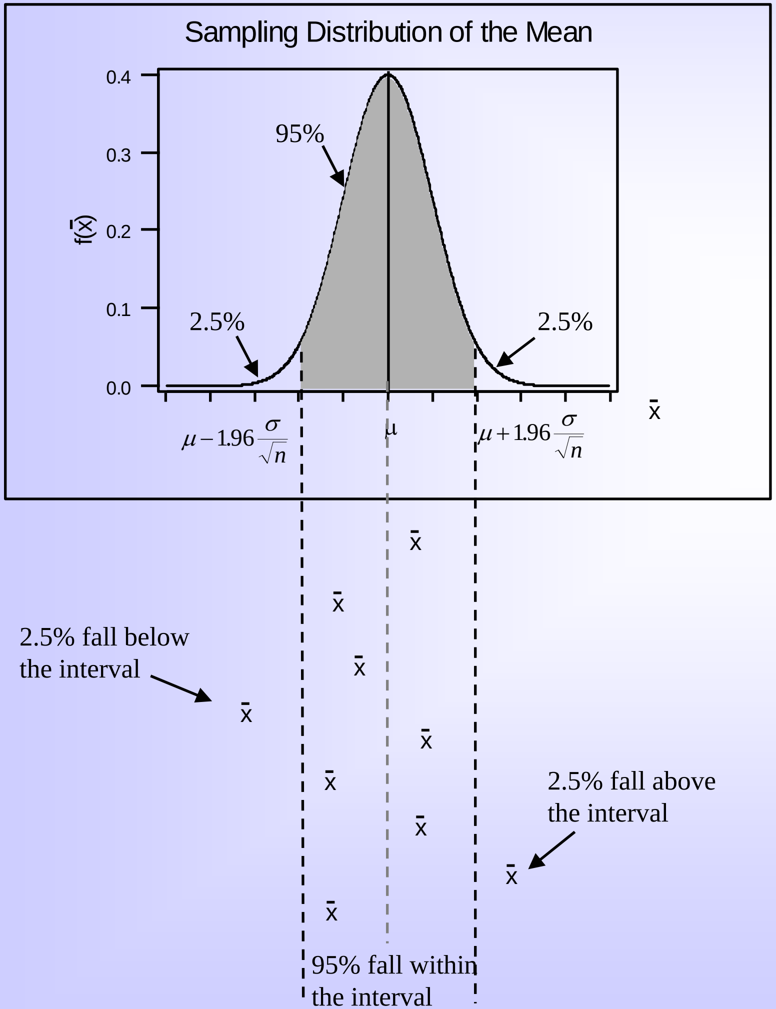
$$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

will include the population mean (and 5% of them will not).

That is, $\bar{\mathbf{x}} \pm \mathbf{1.96}\frac{\boldsymbol{\sigma}}{\sqrt{\mathbf{n}}}$ **is a 95% confidence interval for μ .**

© The McGraw-Hill Companies, Inc., 1999

A 95% Interval around the Population Mean



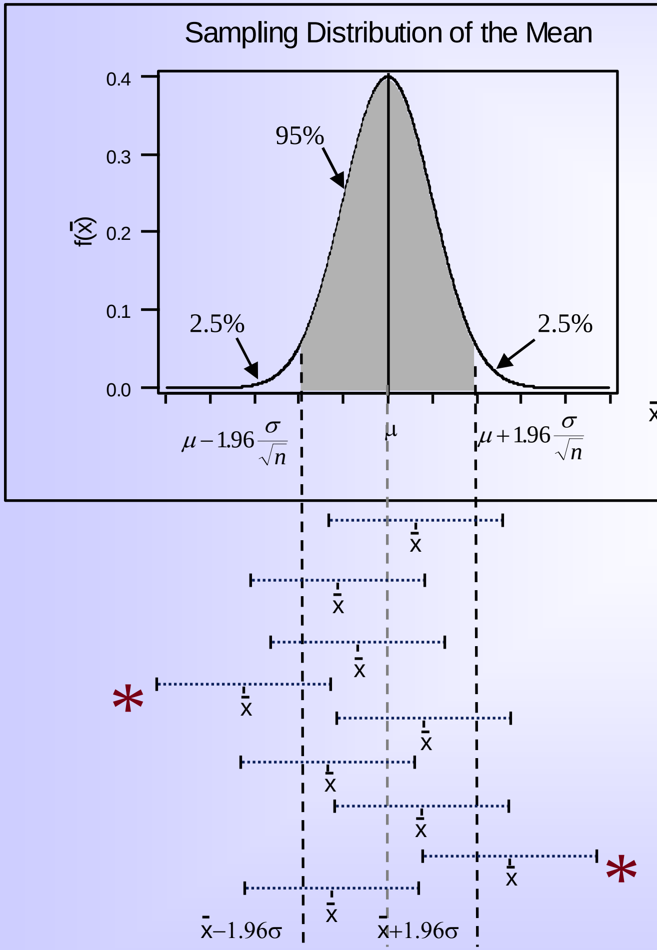
Approximately 95% of sample means can be expected to fall within the interval $\left[\mu - 1.96\frac{\sigma}{\sqrt{n}}, \mu + 1.96\frac{\sigma}{\sqrt{n}}\right]$.

Conversely, about 2.5% can be expected to be above $\mu + 1.96\frac{\sigma}{\sqrt{n}}$ and 2.5% can be expected to be below $\mu - 1.96\frac{\sigma}{\sqrt{n}}$.

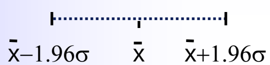
So 5% can be expected to fall outside the interval $\left[\mu - 1.96\frac{\sigma}{\sqrt{n}}, \mu + 1.96\frac{\sigma}{\sqrt{n}}\right]$.

© The McGraw-Hill Companies, Inc., 1999

95% Intervals around the Sample Mean



Approximately 95% of the intervals $x \pm 1.96\frac{\sigma}{\sqrt{n}}$ around the sample mean can be expected to include the actual value of the population mean, μ. (When the sample mean falls within the 95% interval around the population mean.)



*5% of such intervals around the sample mean can be expected *not* to include the actual value of the population mean. (When the sample mean falls outside the 95% interval around the population mean.)

© The McGraw-Hill Companies, Inc., 1999

The 95% Confidence Interval for

A 95% confidence interval for μ when σ is known and sampling is done from a normal population, or a large sample is used:

$$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

The quantity $1.96\frac{\sigma}{\sqrt{n}}$ is often called the *margin of error* or the *sampling error*.

For example, if: n = 25

σ = 20

x = 122

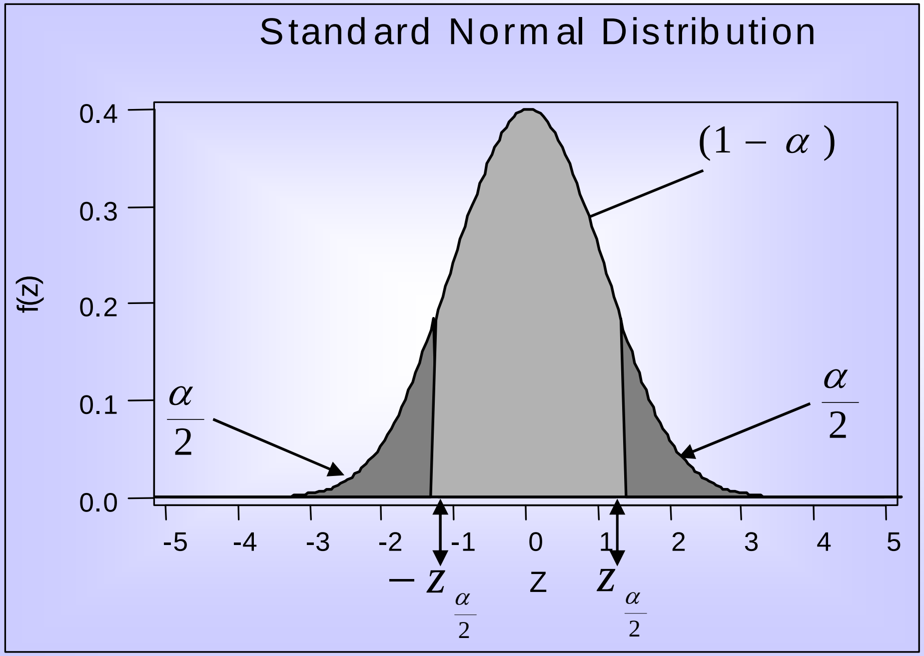
A 95% confidence interval:

$$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}} = 122 \pm 1.96\frac{20}{\sqrt{25}}$$
$$= 122 \pm (1.96)(4)$$
$$= 122 \pm 7.84$$
$$= [114.16, 129.84]$$

© The McGraw-Hill Companies, Inc., 1999

$(1-\alpha)100\%$ Confidence Interval

We define $z_{\frac{\alpha}{2}}$ as the z value that cuts off a right-tail area of $\frac{\alpha}{2}$ under the standard normal curve. (1-α) is called the *confidence coefficient*. α is called the *error probability*, and (1-α)100% is called the *confidence level*.



$$P\left(z > z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

$$P\left(z < -z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

$$P\left(-z_{\frac{\alpha}{2}} < z < z_{\frac{\alpha}{2}}\right) = (1-\alpha)$$

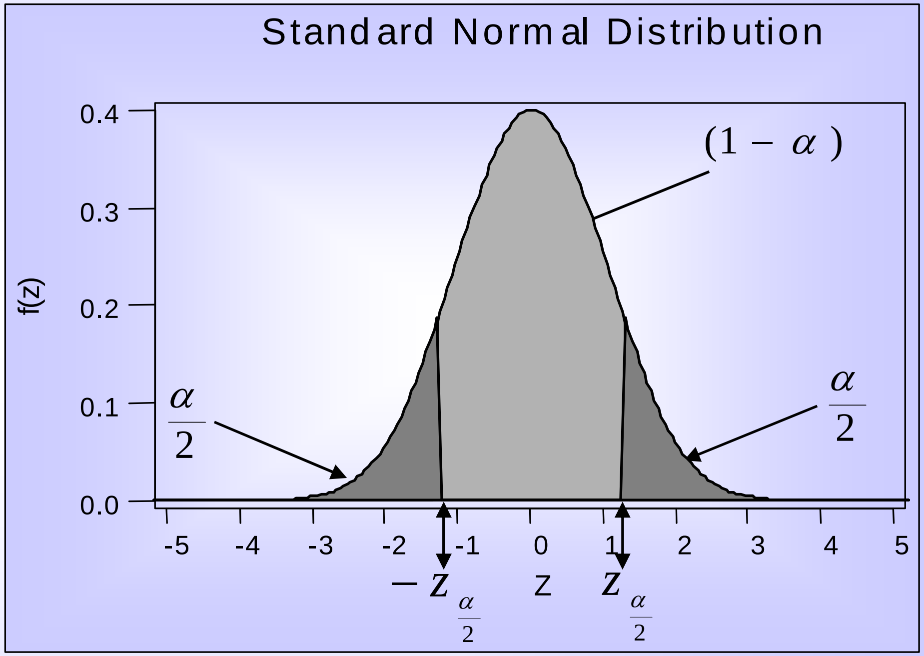
$(1-\alpha)100\%$ Confidence Interval:

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

© The McGraw-Hill Companies, Inc., 1999

Critical Values of z and Levels of Confidence

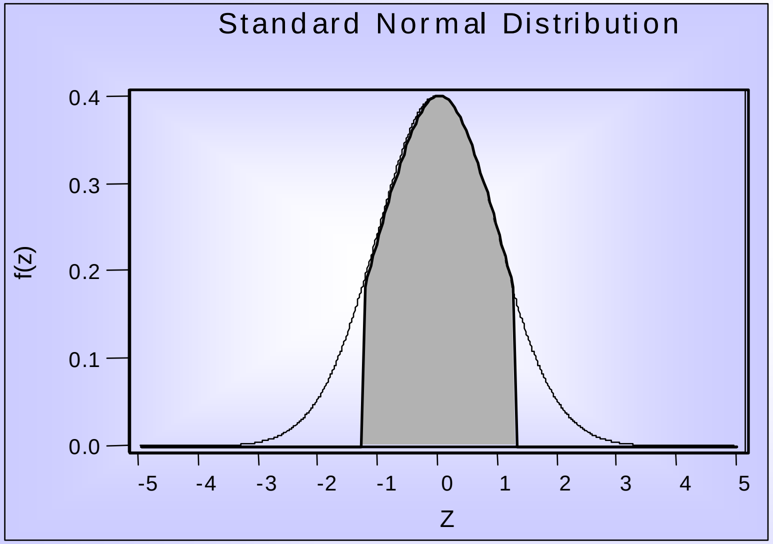
$(1-\alpha)$	$\frac{\alpha}{2}$	$z_{\frac{\alpha}{2}}$
0.99	0.005	2.576
0.98	0.010	2.326
0.95	0.025	1.960
0.90	0.050	1.645
0.80	0.100	1.282



© The McGraw-Hill Companies, Inc., 1999

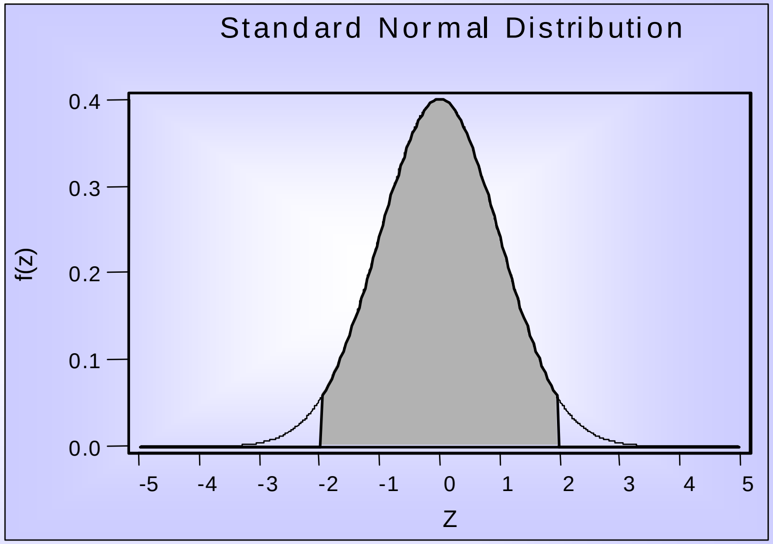
The Level of Confidence and the Width of the Confidence Interval

When sampling from the same population, using a fixed sample size, the *higher the confidence level, the wider the confidence interval.*



80% Confidence Interval:

$$\bar{x} \pm 1.28 \frac{\sigma}{\sqrt{n}}$$



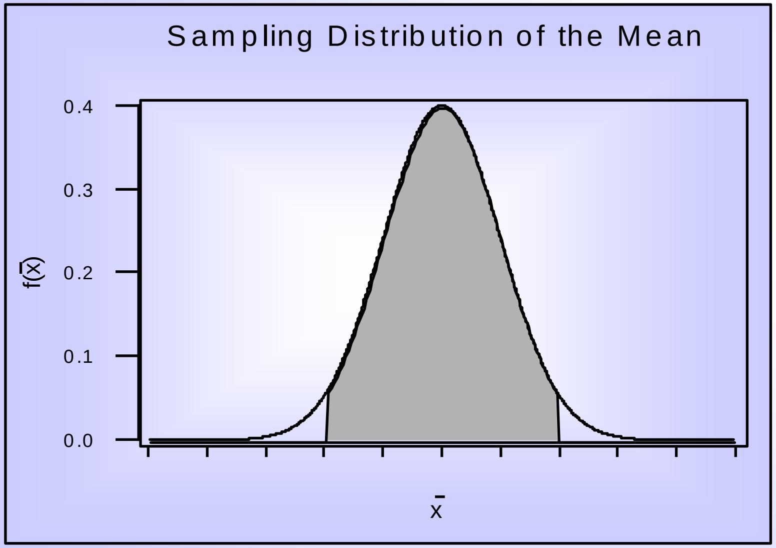
95% Confidence Interval:

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

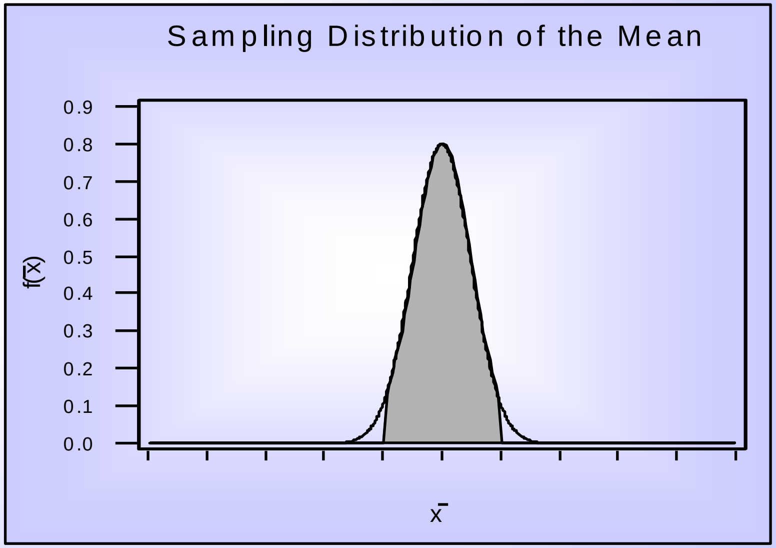
© The McGraw-Hill Companies, Inc., 1999

The Sample Size and the Width of the Confidence Interval

When sampling from the same population, using a fixed confidence level, the *larger the sample size, n, the narrower the confidence interval*.



95% Confidence Interval: n = 20



95% Confidence Interval: n = 40

© The McGraw-Hill Companies, Inc., 1999

Example 6-1

- Population consists of the Fortune 500 Companies (Fortune Web Site), as ranked by Revenues. You are trying to to find out the average Revenues for the companies on the list. The population standard deviation is $15056.37. A random sample of 30 companies obtains a sample mean of $10672.87. Give a 95% and 90% confidence interval for the average Revenues.

© The McGraw-Hill Companies, Inc., 1999

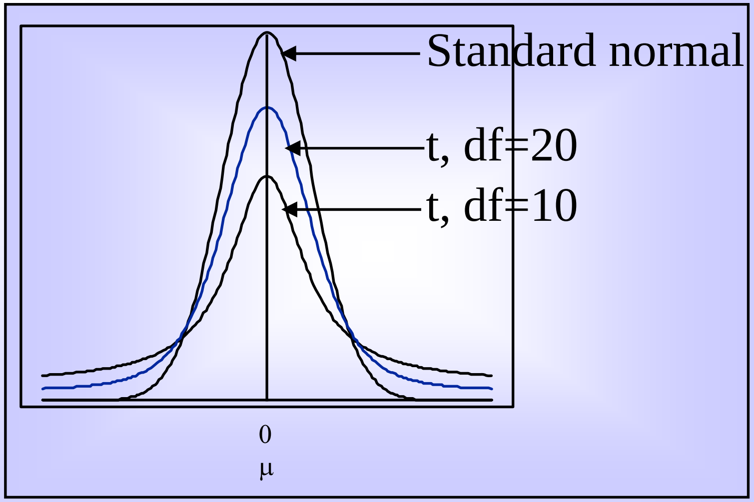
6-3 Confidence Interval or Interval Estimate for μ When σ Is Unknown - The t Distribution

If the population standard deviation, σ, is not known, replace σ with the sample standard deviation, *s*. If the population is normal, the resulting statistic:

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

has a *t distribution* with *(n - 1) degrees of freedom.*

- The *t* is a family of bell-shaped and symmetric distributions, one for each number of degree of freedom.
- The expected value of *t* is 0.
- For *df* > 2, the variance of *t* is *df/(df-2)*. This is greater than 1, but approaches 1 as the number of degrees of freedom increases. The *t* is flatter and has fatter tails than does the standard normal.
- The *t* distribution approaches a standard normal as the number of degrees of freedom increases



© The McGraw-Hill Companies, Inc., 1999

6-3 Confidence Intervals for μ when σ is Unknown- The t Distribution

A *(1-α)*100% confidence interval for μ when σ is not known (assuming a normally distributed population):

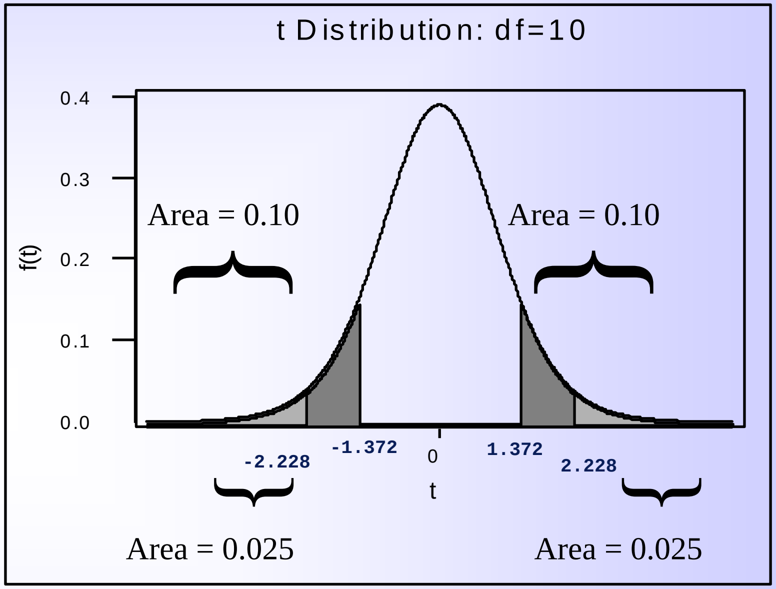
$$\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

where $t_{\frac{\alpha}{2}}$ is the value of the *t* distribution with *n-1* degrees of freedom that cuts off a tail area of $\frac{\alpha}{2}$ to its right.

© The McGraw-Hill Companies, Inc., 1999

The t Distribution

df	$t_{0.100}$	$t_{0.050}$	$t_{0.025}$	$t_{0.010}$	$t_{0.005}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	**1.372**	**1.812**	**2.228**	**2.764**	**3.169**
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	**1.282**	**1.645**	**1.960**	**2.326**	**2.576**



Whenever σ is not known (and the population is assumed normal), the correct distribution to use is the *t* distribution with *n-1* degrees of freedom. Note, however, that for large degrees of freedom, the t distribution is approximated well by the *Z* distribution.

© The McGraw-Hill Companies, Inc., 1999

The t Distribution (Example 6-2)

A stock market analyst wants to estimate the average return on a certain stock. A random sample of 15 days yields an average (annualized) return of $\bar{x}$=10.37% and a standard deviation of s = 3.5%. Assuming a normal population of returns, give a 95% confidence interval for the average return on this stock.

df	$t_{0.100}$	$t_{0.050}$	$t_{0.025}$	$t_{0.010}$	$t_{0.005}$
1	3.078	6.314	12.706	31.821	63.657
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	**2.145**	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

The critical value of t for df = (n-1)=(15-1)=14 and a right-tail area of 0.025 is:

$$t_{0.025} = 2.145$$

The corresponding confidence interval or interval estimate is:

$$\bar{x} \pm t_{0.025}\frac{s}{\sqrt{n}}$$

$$= 10.37 \pm 2.145\frac{3.5}{\sqrt{15}}$$

$$= 10.37 \pm 1.94$$

$$= [8.43, 12.31]$$

© The McGraw-Hill Companies, Inc., 1999

The t Distribution (Example 6-2) Using Excel

Note: The TINV function was used to determine the critical t value.

SAMPLE SIZE	15
SAMPLE MEAN	10.37
SAMPLE STANDARD DEVIATION	3.5
DESIRED CONFIDENCE LEVEL	0.95
ALPHA	0.05
DEGREES OF FREEDOM	14
MEAN OF SAMPLING DISTRIBUTION	10.37
STANDARD ERROR OF SAMPLE MEAN	0.903696
CRITICAL t-VALUE FOR 95% CONFIDENCE INTERVAL	2.144789
HALF-WIDTH OF CONFIDENCE INTERVAL	1.938237
LOWER LIMIT OF CONFIDENCE INTERVAL	8.431763
UPPER LIMIT OF CONFIDENCE INTERVAL	12.30824

© The McGraw-Hill Companies, Inc., 1999

Large Sample Confidence Intervals for the Population Mean

df	$t_{0.100}$	$t_{0.050}$	$t_{0.025}$	$t_{0.010}$	$t_{0.005}$
1	3.078	6.314	12.706	31.821	63.657
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
120	1.289	1.658	1.980	2.358	2.617
∞	**1.282**	**1.645**	**1.960**	**2.326**	**2.576**

Whenever σ is not known (and the population is assumed normal), the correct distribution to use is the *t* distribution with *n-1* degrees of freedom. Note, however, that for large degrees of freedom, the t distribution is approximated well by the *Z* distribution.

© The McGraw-Hill Companies, Inc., 1999

Large Sample Confidence Intervals for the Population Mean

A large - sample $(1-\alpha)100\%$ confidence interval for μ:

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

Example 6-3: An economist wants to estimate the average amount in checking accounts at banks in a given region. A random sample of 100 accounts gives *x-bar* = \$357.60 and *s* = \$140.00. Give a 95% confidence interval for μ, the average amount in any checking account at a bank in the given region.

$$\bar{x} \pm z_{0.025} \frac{s}{\sqrt{n}} = 357.60 \pm 1.96 \frac{140.00}{\sqrt{100}} = 357.60 \pm 27.44 = [330.16, 385.04]$$

© The McGraw-Hill Companies, Inc., 1999

6-4 Large-Sample Confidence Intervals for the Population Proportion, p

The estimator of the population proportion, p, is the sample proportion, $\hat{p}$. If the sample size is large, $\hat{p}$ has an approximately normal distribution, with $E(\hat{p}) = p$ and $V(\hat{p}) = \frac{pq}{n}$, where $q = (1 - p)$. When the population proportion is unknown, use the estimated value, $\hat{p}$, to estimate the standard deviation of $\hat{p}$.

For estimating p, a sample is considered large enough when both $n \cdot p$ an $n \cdot q$ are greater than 5.

© The McGraw-Hill Companies, Inc., 1999

6-4 Large-Sample Confidence Intervals for the Population Proportion, p

A large - sample $(1 - \alpha)100\%$ confidence interval for the population proportion, p:

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

where the sample proportion, $\hat{p}$, is equal to the number of successes in the sample, x, divided by the number of trials (the sample size), n, and $\hat{q} = 1 - \hat{p}$.

© The McGraw-Hill Companies, Inc., 1999

Large-Sample Confidence Interval for the Population Proportion, p (Example 6-4)

A marketing research firm wants to estimate the share that foreign companies have in the American market for certain products. A random sample of 100 consumers is obtained, and it is found that 34 people in the sample are users of foreign-made products; the rest are users of domestic products. Give a 95% confidence interval for the share of foreign products in this market.

$$\begin{aligned}\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} &= 0.34 \pm 1.96\sqrt{\frac{(0.34)(0.66)}{100}} \\ &= 0.34 \pm (1.96)(0.04737) \\ &= 0.34 \pm 0.0928 \\ &= [0.2472, 0.4328]\end{aligned}$$

Thus, the firm may be 95% confident that foreign manufacturers control anywhere from 24.72% to 43.28% of the market.

© The McGraw-Hill Companies, Inc., 1999

Reducing the Width of Confidence Intervals - The Value of Information

The width of a confidence interval can be reduced only at the price of:

- a lower level of confidence, or
- a larger sample.

Lower Level of Confidence

90% Confidence Interval

$$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} = 0.34 \pm 1.645\sqrt{\frac{(0.34)(0.66)}{100}}$$
$$= 0.34 \pm (1.645)(0.04737)$$
$$= 0.34 \pm 0.07792$$
$$= [0.2621, 0.4197]$$

Larger Sample Size

Sample Size, n = 200

$$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} = 0.34 \pm 1.96\sqrt{\frac{(0.34)(0.66)}{200}}$$
$$= 0.34 \pm (1.96)(0.03350)$$
$$= 0.34 \pm 0.0657$$
$$= [0.2743, 0.4057]$$

© The McGraw-Hill Companies, Inc., 1999

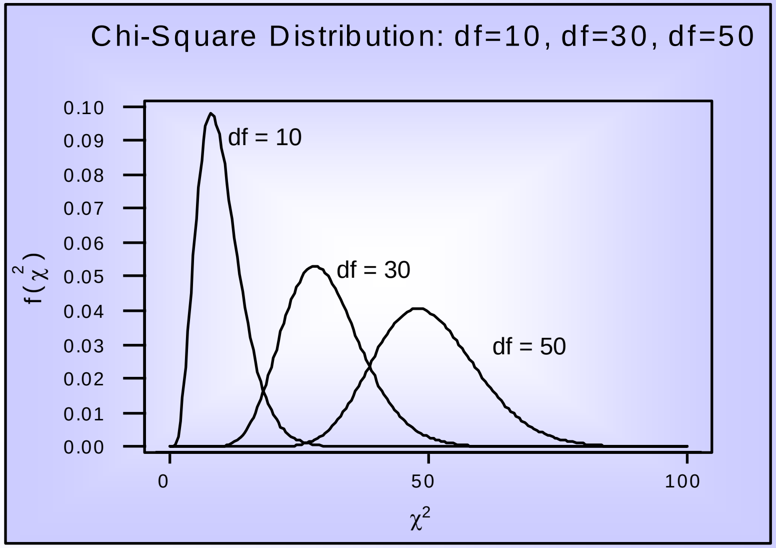
6-5 Confidence Intervals for the Population Variance: The Chi-Square (χ^2) Distribution

- The sample variance, s^2, is an unbiased estimator of the population variance, σ^2.
- Confidence intervals for the population variance are based on the chi-square (χ^2) distribution.
  - The **chi-square distribution** is the probability distribution of the sum of several independent, squared standard normal random variables.
  - The mean of the chi-square distribution is equal to the degrees of freedom parameter, **($E[\chi^2]=df$).** The variance of a chi-square is equal to twice the number of degrees of freedom, **($V[\chi^2]=2df$)**.

© The McGraw-Hill Companies, Inc., 1999

The Chi-Square (χ^2) Distribution

- The chi-square random variable cannot be negative, so it is bound by zero on the left.
- The chi-square distribution is skewed to the right.
- The chi-square distribution approaches a normal as the degrees of freedom increase.



In sampling from a normal population, the random variable:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

has a chi-square distribution with (n - 1) degrees of freedom.

© The McGraw-Hill Companies, Inc., 1999

Values and Probabilities of Chi-Square Distributions

	Area in Right Tail									
	.995	.990	.975	.950	.900	.100	.050	.025	.010	.005
	Area in Left Tail									
df	.005	.010	.025	.050	.100	.900	.950	.975	.990	.995
1	0.0000393	0.000157	0.000982	0.000393	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	4.61	5.99	7.38	9.21	10.60
3	0.0717	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.06	7.78	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	0.676	0.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.95
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.65
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67

© The McGraw-Hill Companies, Inc., 1999

Confidence Interval for the Population Variance

A (1-α)100% confidence interval for the population variance * (where the population is assumed normal):

$$\left[\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}, \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}\right]$$

where $\chi^2_{\frac{\alpha}{2}}$ is the value of the chi-square distribution with n-1 degrees of freedom that cuts off an area $\frac{\alpha}{2}$ to its right and $\chi^2_{1-\frac{\alpha}{2}}$ is the value of the distribution that cuts off an area of $\frac{\alpha}{2}$ to its left (equivalently, an area of $1-\frac{\alpha}{2}$ to its right).

* Note: Because the chi-square distribution is skewed, the confidence interval for the population variance is not symmetric

© The McGraw-Hill Companies, Inc., 1999

Confidence Interval for the Population Variance (Example 6-5)

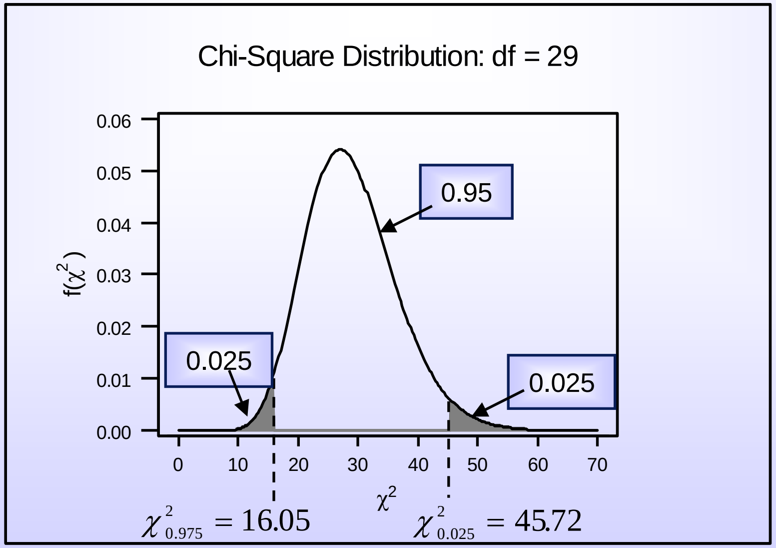
In an automated process, a machine fills cans of coffee. If the average amount filled is different from what it should be, the machine may be adjusted to correct the mean. If the *variance* of the filling process is too high, however, the machine is out of control and needs to be repaired. Therefore, from time to time regular checks of the variance of the filling process are made. This is done by randomly sampling filled cans, measuring their amounts, and computing the sample variance. A random sample of 30 cans gives an estimate $s^2 = 18{,}540$. Give a 95% confidence interval for the population variance, σ^2.

$$\left[\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}, \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}\right] = \left[\frac{(30-1)18540}{45.7}, \frac{(30-1)18540}{16.0}\right] = [11765, 33604]$$

© The McGraw-Hill Companies, Inc., 1999

Example 6-5 (continued)

	Area in Right Tail									
df	.995	.990	.975	.950	.900	.100	.050	.025	.010	.005
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67



© The McGraw-Hill Companies, Inc., 1999

6-6 Sample-Size Determination

Before determining the necessary sample size, three questions must be answered:

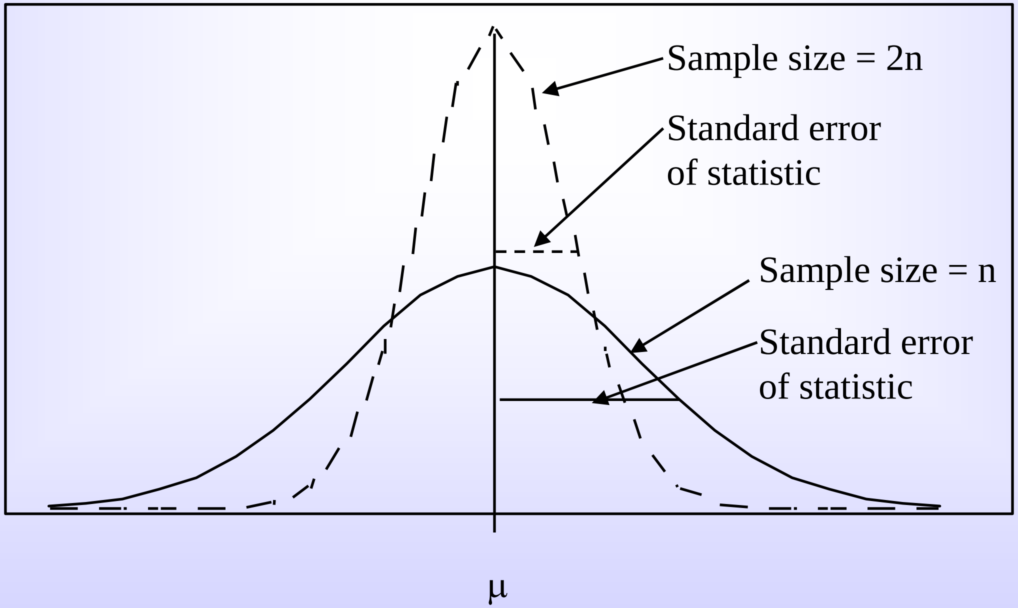
- How close do you want your sample estimate to be to the unknown parameter? (What is the desired ***bound***, **B**?)
- What do you want the desired confidence level **(1-α)** to be so that the distance between your estimate and the parameter is less than or equal to B?
- What is your estimate of the variance (or standard deviation) of the population in question?

For example: A $(1-\alpha)$ Confidence Interval for μ: $\bar{x} \pm \underbrace{z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}_{\text{Bound, B}}$

© The McGraw-Hill Companies, Inc., 1999

Sample Size and Standard Error

The sample size determines the bound of a statistic, since the standard error of a statistic shrinks as the sample size increases:



© The McGraw-Hill Companies, Inc., 1999

Minimum Sample Size: Mean and Proportion

Minimum required sample size in estimating the population mean, μ:

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{B^2}$$

Bound of estimate:

$$B = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Minimum required sample size in estimating the population proportion, p

$$n = \frac{z_{\alpha/2}^2 pq}{B^2}$$

© The McGraw-Hill Companies, Inc., 1999

Sample-Size Determination (Example 6-6)

A marketing research firm wants to conduct a survey to estimate the average amount spent on entertainment by each person visiting a popular resort. The people who plan the survey would like to determine the average amount spent by all people visiting the resort to within \$120, with 95% confidence. From past operation of the resort, an estimate of the population standard deviation is s = \$400. What is the minimum required sample size?

$$n = \frac{z^2_{\frac{\alpha}{2}}\sigma^2}{B^2}$$

$$= \frac{(1.96)^2(400)^2}{120^2}$$

$$= 42.684 \approx 43$$

© The McGraw-Hill Companies, Inc., 1999

Sample-Size for Proportion (Example 6-7)

The manufacturers of a sports car want to estimate the proportion of people in a given income bracket who are interested in the model. The company wants to know the population proportion, *p*, to within 0.01 with 99% confidence. Current company records indicate that the proportion *p* may be around 0.25. What is the minimum required sample size for this survey?

$$n = \frac{z^2_{\frac{\alpha}{2}} pq}{B^2}$$

$$= \frac{2.576^2 (0.25)(0.75)}{0.10^2}$$

$$= 124.42 \approx 125$$

© The McGraw-Hill Companies, Inc., 1999

Sample-Size for Proportion (Example 6-7) Using Excel

The required sample sizes for various degrees of accuracy (express as interval half-widths) are displayed for 90%, 95%, and 99% confidence levels are shown on the next slide. The specific case requested in the example, where the desired half-width is $120 with a 95% confidence level, appears in the bordered cell. The NORMINV function was used to calculate the critical z values for each confidence interval and the ROUNDUP function was used to get the smallest integer value that is at least as great as the required sample size determined by the formulas in the text

© The McGraw-Hill Companies, Inc., 1999

Sample-Size for Proportion (Example 6-7) Using Excel

DESIRED ACCURACY HALF-WIDTH)	REQUIRE SAMPLE SIZE FOR 90% CONFIDENCE	FOR 95% CONFIDENCE	FOR 99% CONFIDENC
20	1083	1537	2654
40	271	385	664
60	121	171	295
80	68	97	166
100	44	62	107
120	31	$\boxed{43}$	74
140	23	32	55
160	17	25	42
180	14	19	33
200	11	16	27
220	9	13	22
240	8	11	19
260	7	10	16
280	6	8	14
300	5	7	12
320	5	7	11
340	4	6	10
360	4	5	9
380	3	5	8

© The McGraw-Hill Companies, Inc., 1999

6-7 Using the Computer: Confidence Intervals

```
MTB > Set C1
DATA> 17.125
DATA> 17
DATA> 17.375
DATA> 17.25
DATA> 16.625
DATA> 16.875
DATA> 17
DATA> 17
DATA> 17.25
DATA> 18
DATA> 18.125
DATA> 17.5
DATA> 17.375
DATA> 17.5
DATA> 17.25
DATA> end
```

```
MTB > tinterval c1

Confidence Intervals


Variable     N      Mean    StDev  SE Mean       95.0 % C.I.
C1          15    17.283    0.397    0.102  (  17.064,  17.503)

MTB > zinterval 0.397 c1

Confidence Intervals


The assumed sigma = 0.397

Variable     N      Mean    StDev  SE Mean       95.0 % C.I.
C1          15    17.283    0.397    0.103  (  17.082,  17.484)
```


© The McGraw-Hill Companies, Inc., 1999

Using the Computer: t and Chi-Square Probabilities

```
MTB > CDF 2.00;
SUBC> t df=24.

Cumulative Distribution Function

Student's t distribution with 24 d.f.

        x      P( X <= x)
    2.0000        0.9715

MTB > INVCDF 0.95;
SUBC> T df=24.

Inverse Cumulative Distribution Function

Student's t distribution with 24 d.f.

 P( X <= x)          x
    0.9500        1.7109
```

```
MTB > CDF 35;
SUBC> CHISQUARE 25.

Cumulative Distribution Function

Chisquare with 25 d.f.

        x      P( X <= x)
   35.0000        0.9118

MTB > INVCDF 0.99;
SUBC> CHISQUARE 20.

Inverse Cumulative Distribution Function

Chisquare with 20 d.f.

 P( X <= x)          x
    0.9900       37.5662
```


© The McGraw-Hill Companies, Inc., 1999

Using the Computer: Analysis of Data in Fig. 6-11 Using Excel

The DESCRIPTIVE STATISTICS option in the DATA ANALYSIS TOOLKIT (found in the TOOLS menu if the toolkit has been added) was used to describe the data, which appears on the left of the screen (see the next slide). The value appearing next to "Confidence Level (95%)" is the half-width for the 95% confidence interval on the mean. Adding or subtracting this value from the sample mean gives the upper and lower limits of the confidence interval, respectively.

© The McGraw-Hill Companies, Inc., 1999

Using the Computer: Analysis of Data in Fig. 6-11 Using Excel

DATA VALUES
Stock Price
17.125
17
17.375
17.25
16.625
16.875
17
17
17.25
18
18.125
17.5
17.375
17.5
17.25

	Stock Price
Mean	17.28333333
Standard Error	0.102401714
Median	17.25
Mode	17
Standard Deviation	0.396600134
Sample Variance	0.157291667
Kurtosis	0.698140288
Skewness	0.740501862
Range	1.5
Minimum	16.625
Maximum	18.125
Sum	259.25
Count	15
Confidence Level(95.0%)	0.219630029
LOWER LIMIT	17.0637033
UPPER LIMIT	17.50296336

© The McGraw-Hill Companies, Inc., 1999