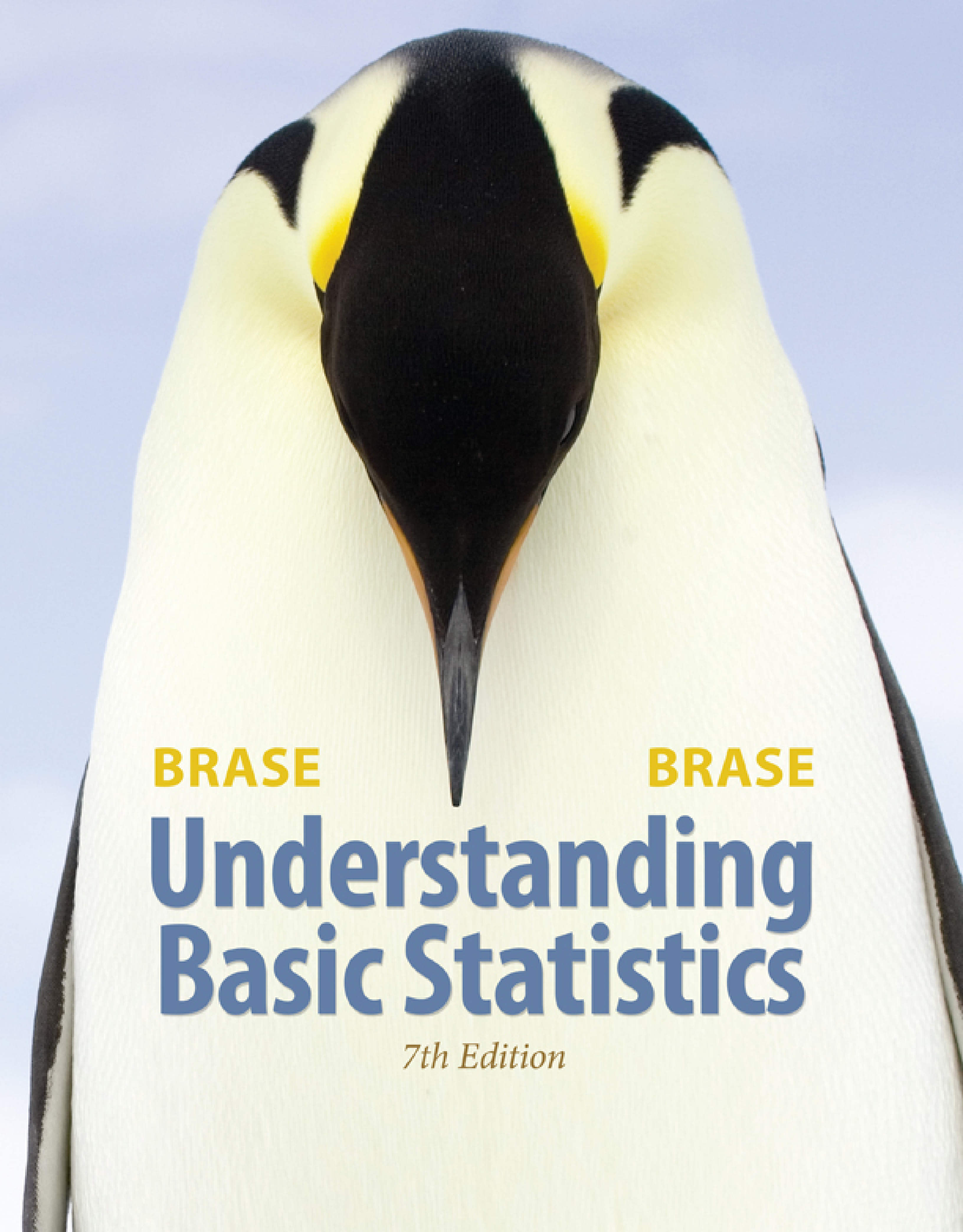


BRASE BRASE

Understanding Basic Statistics

7th Edition

FREQUENTLY USED FORMULAS

n = sample size $\quad N$ = population size $\quad f$ = frequency

Chapter 2

Class width $= \dfrac{\text{high} - \text{low}}{\text{number of classes}}$ (increase to next integer)

Class midpoint $= \dfrac{\text{upper limit} + \text{lower limit}}{2}$

Lower boundary = lower boundary of previous class + class width

Chapter 3

Sample mean $\bar{x} = \dfrac{\Sigma x}{n}$

Population mean $\mu = \dfrac{\Sigma x}{N}$

Weighted average $= \dfrac{\Sigma xw}{\Sigma w}$

Range = largest data value − smallest data value

Sample standard deviation $s = \sqrt{\dfrac{\Sigma(x - \bar{x})^2}{n - 1}}$

Computation formula $s = \sqrt{\dfrac{\Sigma x^2 - (\Sigma x)^2/n}{n - 1}}$

Population standard deviation $\sigma = \sqrt{\dfrac{\Sigma(x - \mu)^2}{N}}$

Sample variance s^2

Population variance σ^2

Sample coefficient of variation $CV = \dfrac{s}{\bar{x}} \cdot 100\%$

Sample mean for grouped data $\bar{x} = \dfrac{\Sigma xf}{n}$

Sample standard deviation for grouped data

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2 f}{n - 1}} = \sqrt{\frac{\Sigma x^2 f - (\Sigma xf)^2/n}{n - 1}}$$

Chapter 4

Regression and Correlation

Pearson product moment correlation coefficient

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

Least-squares line $\hat{y} = a + bx$

where $b = \dfrac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$

$a = \bar{y} - b\bar{x}$

Coefficient of determination $= r^2$

Chapter 5

Probability of the complement of event A

$P(A^c) = 1 - P(A)$

Multiplication rule for independent events

$P(A \text{ and } B) = P(A) \cdot P(B)$

General multiplication rules

$P(A \text{ and } B) = P(A) \cdot P(B|A)$

$P(A \text{ and } B) = P(B) \cdot P(A|B)$

Addition rule for mutually exclusive events

$P(A \text{ or } B) = P(A) + P(B)$

General addition rule

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

Permutation rule $P_{n,r} = \dfrac{n!}{(n - r)!}$

Combination rule $C_{n,r} = \dfrac{n!}{r!(n - r)!}$

Chapter 6

Mean of a discrete probability distribution $\mu = \Sigma xP(x)$

Standard deviation of a discrete probability distribution

$\sigma = \sqrt{\Sigma(x - \mu)^2 P(x)}$

For Binomial Distributions

r = number of successes; p = probability of success; $q = 1 - p$

Binomial probability distribution $P(r) = C_{n,r}p^r q^{n-r}$

Mean $\mu = np$

Standard deviation $\sigma = \sqrt{npq}$

Chapter 7

Raw score $x = z\sigma + \mu$ $\qquad$ Standard score $z = \dfrac{x - \mu}{\sigma}$

Mean of $\bar{x}$ distribution $\mu_{\bar{x}} = \mu$

Standard deviation of $\bar{x}$ distribution $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

Standard score for $\bar{x}$; $\quad z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

Mean of $\hat{p}$ distribution $\mu_{\hat{p}} = p$

Standard deviation for $\hat{p}$; $\quad \sigma_{\hat{p}} = \sqrt{\dfrac{pq}{n}}$; $q = 1 - p$

Chapter 8

Confidence Interval

for μ

$\bar{x} - E < \mu < \bar{x} + E$

where $E = z_c \dfrac{\sigma}{\sqrt{n}}$ when σ is known

FREQUENTLY USED FORMULAS

n = sample size N = population size f = frequency

Chapter 2

Class width $= \dfrac{\text{high} - \text{low}}{\text{number of classes}}$ (increase to next integer)

Class midpoint $= \dfrac{\text{upper limit} + \text{lower limit}}{2}$

Lower boundary = lower boundary of previous class + class width

Chapter 3

Sample mean $\bar{x} = \dfrac{\Sigma x}{n}$

Population mean $\mu = \dfrac{\Sigma x}{N}$

Weighted average $= \dfrac{\Sigma xw}{\Sigma w}$

Range = largest data value − smallest data value

Sample standard deviation $s = \sqrt{\dfrac{\Sigma(x - \bar{x})^2}{n - 1}}$

Computation formula $s = \sqrt{\dfrac{\Sigma x^2 - (\Sigma x)^2/n}{n - 1}}$

Population standard deviation $\sigma = \sqrt{\dfrac{\Sigma(x - \mu)^2}{N}}$

Sample variance s^2

Population variance σ^2

Sample coefficient of variation $CV = \dfrac{s}{\bar{x}} \cdot 100\%$

Sample mean for grouped data $\bar{x} = \dfrac{\Sigma xf}{n}$

Sample standard deviation for grouped data

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2 f}{n - 1}} = \sqrt{\frac{\Sigma x^2 f - (\Sigma xf)^2/n}{n - 1}}$$

Chapter 4

Regression and Correlation

Pearson product moment correlation coefficient

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

Least-squares line $\hat{y} = a + bx$

where $b = \dfrac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$

$a = \bar{y} - b\bar{x}$

Coefficient of determination $= r^2$

Chapter 5

Probability of the complement of event A

$P(A^c) = 1 - P(A)$

Multiplication rule for independent events

$P(A \text{ and } B) = P(A) \cdot P(B)$

General multiplication rules

$P(A \text{ and } B) = P(A) \cdot P(B|A)$

$P(A \text{ and } B) = P(B) \cdot P(A|B)$

Addition rule for mutually exclusive events

$P(A \text{ or } B) = P(A) + P(B)$

General addition rule

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

Permutation rule $P_{n,r} = \dfrac{n!}{(n - r)!}$

Combination rule $C_{n,r} = \dfrac{n!}{r!(n - r)!}$

Chapter 6

Mean of a discrete probability distribution $\mu = \Sigma xP(x)$

Standard deviation of a discrete probability distribution

$\sigma = \sqrt{\Sigma(x - \mu)^2 P(x)}$

For Binomial Distributions

r = number of successes; p = probability of success; $q = 1 - p$

Binomial probability distribution $P(r) = C_{n,r}p^r q^{n-r}$

Mean $\mu = np$

Standard deviation $\sigma = \sqrt{npq}$

Chapter 7

Raw score $x = z\sigma + \mu$ Standard score $z = \dfrac{x - \mu}{\sigma}$

Mean of $\bar{x}$ distribution $\mu_{\bar{x}} = \mu$

Standard deviation of $\bar{x}$ distribution $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

Standard score for $\bar{x}$; $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

Mean of $\hat{p}$ distribution $\mu_{\hat{p}} = p$

Standard deviation for $\hat{p}$; $\sigma_{\hat{p}} = \sqrt{\dfrac{pq}{n}}$; $q = 1 - p$

Chapter 8

Confidence Interval

for μ

$\bar{x} - E < \mu < \bar{x} + E$

where $E = z_c \dfrac{\sigma}{\sqrt{n}}$ when σ is known

$E = t_c \dfrac{s}{\sqrt{n}}$ when σ is unknown

with $d.f. = n - 1$

for p $(np > 5$ and $n(1 - p) > 5)$

$\hat{p} - E < p < \hat{p} + E$

where $E = z_c\sqrt{\dfrac{\hat{p}(1 - \hat{p})}{n}}$

$\hat{p} = \dfrac{r}{n}$

Sample Size for Estimating

means $n = \left(\dfrac{z_c\sigma}{E}\right)^2$

proportions

$n = p(1 - p)\left(\dfrac{z_c}{E}\right)^2$ with preliminary estimate for p

$n = \dfrac{1}{4}\left(\dfrac{z_c}{E}\right)^2$ without preliminary estimate for p

Chapter 9

Sample Test Statistics for Tests of Hypotheses

for μ (σ known) $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

for μ (σ unknown) $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$; $d.f. = n - 1$

for p ($np > 5$ and $nq > 5$) $z = \dfrac{\hat{p} - p}{\sqrt{pq/n}}$

where $q = 1 - p$; $\hat{p} = r/n$

Chapter 10

Sample Test Statistics for Tests of Hypotheses

for paired differences d $t = \dfrac{\bar{d} - \mu_d}{s_d/\sqrt{n}}$; $d.f. = n - 1$

for difference of means, σ_1 and σ_2 known

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

for difference of means, σ_1 or σ_2 unknown

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

$d.f.$ = smaller of $n_1 - 1$ and $n_2 - 1$

(*Note:* Software uses Satterthwaite's approximation for degrees of freedom $d.f.$)

for difference of proportions

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\bar{p}\bar{q}}{n_1} + \dfrac{\bar{p}\bar{q}}{n_2}}}$$

where $\bar{p} = \dfrac{r_1 + r_2}{n_1 + n_2}$ and $\bar{q} = 1 - \bar{p}$

$\hat{p}_1 = r_1/n_1$; $\hat{p}_2 = r_2/n_2$

Confidence Interval

for $\mu_1 - \mu_2$ (independent samples)

$(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E$

where $E = z_c\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$ when σ_1 and σ_2 are known

$E = t_c\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$ when σ_1 or σ_2 is unknown

with $d.f.$ = smaller of $n_1 - 1$ and $n_2 - 1$

(*Note:* Software uses Satterthwaite's approximation for degrees of freedom $d.f.$)

for difference of proportions $p_1 - p_2$

$(\hat{p}_1 - \hat{p}_2) - E < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + E$

where $E = z_c\sqrt{\dfrac{\hat{p}_1\hat{q}_1}{n_1} + \dfrac{\hat{p}_2\hat{q}_2}{n_2}}$

$\hat{p}_1 = r_1/n_1$; $\hat{p}_2 = r_2/n_2$

$\hat{q}_1 = 1 - \hat{p}_1$; $\hat{q}_2 = 1 - \hat{p}_2$

Chapter 11

$\chi^2 = \sum \dfrac{(O - E)^2}{E}$ where O = observed frequency

For tests of independence and tests of homogeneity

$E = \dfrac{\text{(row total)(column total)}}{\text{sample size}}$

For goodness of fit test E = (given percent)(sample size)

Tests of independence $d.f. = (R - 1)(C - 1)$

Test of homogeneity $d.f. = (R - 1)(C - 1)$

Goodness of fit $d.f.$ = (number of categories) − 1

Sample test statistic for σ^2

$\chi^2 = \dfrac{(n - 1)s^2}{\sigma^2}$ with $d.f. = n - 1$

Linear Regression

Sample test statistic for r

$t = \dfrac{r\sqrt{n - 2}}{\sqrt{1 - r^2}}$ with $d.f. = n - 2$

Standard error of estimate $S_e = \sqrt{\dfrac{\Sigma y^2 - a\Sigma y - b\Sigma xy}{n - 2}}$

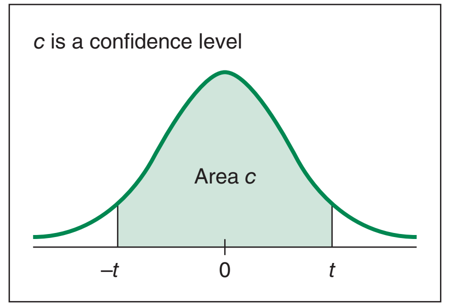
Confidence interval for y: $\hat{y} - E < y < \hat{y} + E$

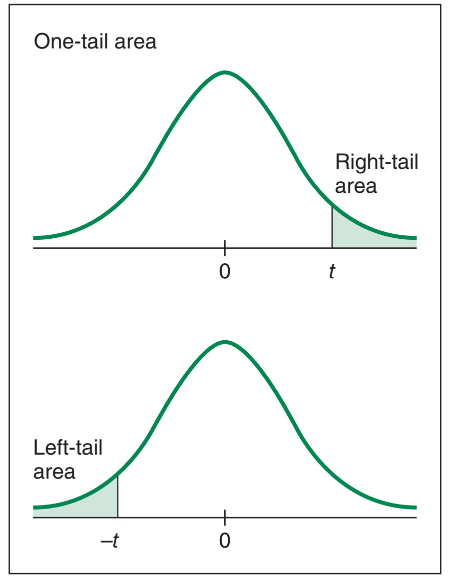
where $E = t_c S_e\sqrt{1 + \dfrac{1}{n} + \dfrac{n(x - \bar{x})^2}{n\Sigma x^2 - (\Sigma x)^2}}$

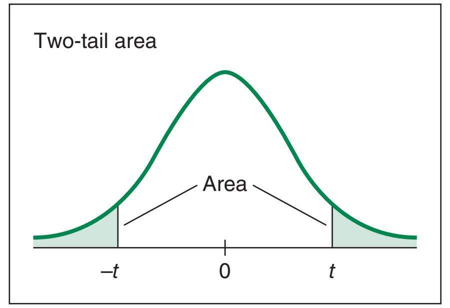
with $d.f. = n - 2$

Sample test statistic for slope b

$t = \dfrac{b}{S_e}\sqrt{\Sigma x^2 - \dfrac{1}{n}(\Sigma x)^2}$ with $d.f. = n - 2$







Critical Values for Student's *t* Distribution

one-tail area	0.250	0.125	0.100	0.075	0.050	0.025	0.010	0.005	0.0005
two-tail area	0.500	0.250	0.200	0.150	0.100	0.050	0.020	0.010	0.0010
d.f. \ *c*	0.500	0.750	0.800	0.850	0.900	0.950	0.980	0.990	0.999
1	1.000	2.414	3.078	4.165	6.314	12.706	31.821	63.657	636.619
2	0.816	1.604	1.886	2.282	2.920	4.303	6.965	9.925	31.599
3	0.765	1.423	1.638	1.924	2.353	3.182	4.541	5.841	12.924
4	0.741	1.344	1.533	1.778	2.132	2.776	3.747	4.604	8.610
5	0.727	1.301	1.476	1.699	2.015	2.571	3.365	4.032	6.869
6	0.718	1.273	1.440	1.650	1.943	2.447	3.143	3.707	5.959
7	0.711	1.254	1.415	1.617	1.895	2.365	2.998	3.499	5.408
8	0.706	1.240	1.397	1.592	1.860	2.306	2.896	3.355	5.041
9	0.703	1.230	1.383	1.574	1.833	2.262	2.821	3.250	4.781
10	0.700	1.221	1.372	1.559	1.812	2.228	2.764	3.169	4.587
11	0.697	1.214	1.363	1.548	1.796	2.201	2.718	3.106	4.437
12	0.695	1.209	1.356	1.538	1.782	2.179	2.681	3.055	4.318
13	0.694	1.204	1.350	1.530	1.771	2.160	2.650	3.012	4.221
14	0.692	1.200	1.345	1.523	1.761	2.145	2.624	2.977	4.140
15	0.691	1.197	1.341	1.517	1.753	2.131	2.602	2.947	4.073
16	0.690	1.194	1.337	1.512	1.746	2.120	2.583	2.921	4.015
17	0.689	1.191	1.333	1.508	1.740	2.110	2.567	2.898	3.965
18	0.688	1.189	1.330	1.504	1.734	2.101	2.552	2.878	3.922
19	0.688	1.187	1.328	1.500	1.729	2.093	2.539	2.861	3.883
20	0.687	1.185	1.325	1.497	1.725	2.086	2.528	2.845	3.850
21	0.686	1.183	1.323	1.494	1.721	2.080	2.518	2.831	3.819
22	0.686	1.182	1.321	1.492	1.717	2.074	2.508	2.819	3.792
23	0.685	1.180	1.319	1.489	1.714	2.069	2.500	2.807	3.768
24	0.685	1.179	1.318	1.487	1.711	2.064	2.492	2.797	3.745
25	0.684	1.198	1.316	1.485	1.708	2.060	2.485	2.787	3.725
26	0.684	1.177	1.315	1.483	1.706	2.056	2.479	2.779	3.707
27	0.684	1.176	1.314	1.482	1.703	2.052	2.473	2.771	3.690
28	0.683	1.175	1.313	1.480	1.701	2.048	2.467	2.763	3.674
29	0.683	1.174	1.311	1.479	1.699	2.045	2.462	2.756	3.659
30	0.683	1.173	1.310	1.477	1.697	2.042	2.457	2.750	3.646
35	0.682	1.170	1.306	1.472	1.690	2.030	2.438	2.724	3.591
40	0.681	1.167	1.303	1.468	1.684	2.021	2.423	2.704	3.551
45	0.680	1.165	1.301	1.465	1.679	2.014	2.412	2.690	3.520
50	0.679	1.164	1.299	1.462	1.676	2.009	2.403	2.678	3.496
60	0.679	1.162	1.296	1.458	1.671	2.000	2.390	2.660	3.460
70	0.678	1.160	1.294	1.456	1.667	1.994	2.381	2.648	3.435
80	0.678	1.159	1.292	1.453	1.664	1.990	2.374	2.639	3.416
100	0.677	1.157	1.290	1.451	1.660	1.984	2.364	2.626	3.390
500	0.675	1.152	1.283	1.442	1.648	1.965	2.334	2.586	3.310
1000	0.675	1.151	1.282	1.441	1.646	1.962	2.330	2.581	3.300
∞	**0.674**	**1.150**	**1.282**	**1.440**	**1.645**	**1.960**	**2.326**	**2.576**	**3.291**

For degrees of freedom (*d.f.*) not in the table, use the closest *d.f.* that is *smaller.*

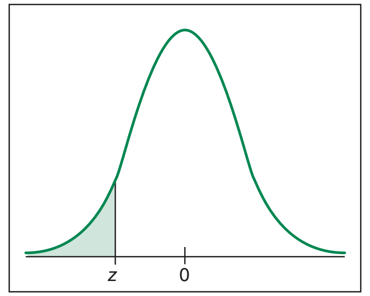


Table entry for z is the area to the left of z.

Areas of a Standard Normal Distribution

(a) Table of Areas to the Left of z

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	.0018	.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
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

For values of z less than −3.49, use 0.000 to approximate the area.

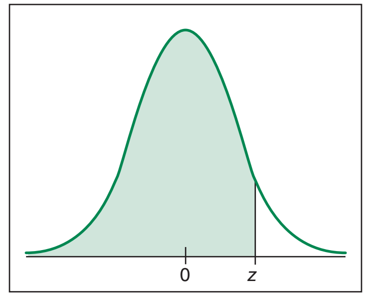


Table entry for z is the area to the left of z.

Areas of a Standard Normal Distribution *continued*

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

For z values greater than 3.49, use 1.000 to approximate the area.

Areas of a Standard Normal Distribution *continued*

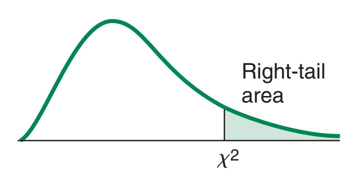
(b) Confidence Interval Critical Values z_c

Level of Confidence c	Critical Value z_c
0.70, or 70%	1.04
0.75, or 75%	1.15
0.80, or 80%	1.28
0.85, or 85%	1.44
0.90, or 90%	1.645
0.95, or 95%	1.96
0.98, or 98%	2.33
0.99, or 99%	2.58

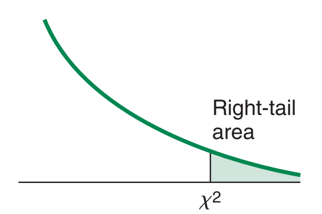
Areas of a Standard Normal Distribution *continued*

(c) Hypothesis Testing, Critical Values z_0

Level of Significance	$\alpha = 0.05$	$\alpha = 0.01$
Critical value z_0 for a left-tailed test	−1.645	−2.33
Critical value z_0 for a right-tailed test	1.645	2.33
Critical values $\pm z_0$ for a two-tailed test	±1.96	±2.58



For $d.f. \geq 3$



For $d.f.$ = 1 or 2

The χ^2 Distribution

$d.f.$	Right-Tail Area .995	.990	.975	.950	.900	.100	.050	.025	.010	.005
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	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	0.072	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	1.064	7.78	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	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	21.96
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	8.59	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	49.64
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.21	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
40	20.71	22.16	24.43	26.51	29.05	51.80	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.4	104.2
80	51.17	53.54	57.15	60.39	64.28	96.58	101.9	106.6	112.3	116.3
90	59.20	61.75	65.65	69.13	73.29	107.6	113.1	118.1	124.1	128.3
100	67.33	70.06	74.22	77.93	82.36	118.5	124.3	129.6	135.8	140.2

Source: From H. L. Herter, *Biometrika,* June 1964. Printed by permission of the Biometrika Trustees.

0.0^4393 means that there are 4 0's between the decimal and 3; $0.0^4393 = 0.0000393$.

UNDERSTANDING BASIC STATISTICS

SEVENTH EDITION

UNDERSTANDING BASIC STATISTICS

Charles Henry Brase
Regis University

Corrinne Pellillo Brase
Arapahoe Community College

CENGAGE Learning®

Australia • Brazil • Mexico • Singapore • United Kingdom • United States

This book is dedicated to the memory of
a great teacher, mathematician and friend

Burton W. Jones
Professor Emeritus, University of Colorado

CENGAGE Learning®

Understanding Basic Statistics,
Seventh Edition
Brase/Brase

Senior Product Team Manager : Richard Stratton

Senior Product Manager: Molly Taylor

Senior Content Developer: Jay Campbell

Associate Content Developer: Danielle Hallock

Product Assistant: Spencer Arritt

Media Developer: Andrew Coppola

Marketing Manager: Ryan Ahern

Content Project Manager: Jill Quinn

Senior Art Director: Linda May

Manufacturing Planner: Sandee Milewski

IP Analyst: Christina Ciaramella

IP Project Manager: Farah Fard

Production Service: Integra Software Services

Compositor: Integra Software Services

Text and Cover Designer: Rokusek Design

Cover and Interior Image: Shutterstock, © Jan Martin Will

Library of Congress Control Number: 2014950204

Student Edition:

ISBN: 978-1-305-25406-0

Annotated Instructor's Edition:

ISBN: 978-1-305-25878-5

Cengage Learning
20 Channel Center Street
Boston, MA 02210
USA

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil, and Japan. Locate your local office at **www.cengage.com/global.**

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

To learn more about Cengage Learning Solutions, visit **www.cengage.com.**

Purchase any of our products at your local college store or at our preferred online store **www.cengagebrain.com.**

Printed in the United States of America
Print Number: 01 Print Year: 2014

CONTENTS

CRITICAL THINKING

Students need to develop critical thinking skills in order to understand and evaluate the limitations of statistical methods. *Understanding Basic Statistics* makes students aware of method appropriateness, assumptions, biases, and justifiable conclusions.

CRITICAL THINKING

UNUSUAL VALUES

Chebyshev's theorem tells us that no matter what the data distribution looks like, at least 75% of the data will fall within 2 standard deviations of the mean. As we will see in Chapter 6, when the distribution is mound-shaped and symmetrical, about 95% of the data are within 2 standard deviations of the mean. Data values beyond 2 standard deviations from the mean are less common than those closer to the mean.

In fact, one indicator that a data value might be an outlier is that it is more than 2.5 standard deviations from the mean (Source: *Statistics,* by G. Upton and I. Cook, Oxford University Press).

UNUSUAL VALUES

For a binomial distribution, it is unusual for the number of successes r to be higher than $\mu + 2.5\sigma$ or lower than $\mu - 2.5\sigma$.

We can use this indicator to determine whether a specified number of successes out of n trials in a binomial experiment are unusual.

For instance, consider a binomial experiment with 20 trials for which probability of success on a single trial is $p = 0.70$. The expected number of successes is $\mu = 14$, with a standard deviation of $\sigma \approx 2$. A number of successes above 19 or below 9 would be considered unusual. However, such numbers of successes are possible.

◀ Critical Thinking

Critical thinking is an important skill for students to develop in order to avoid reaching misleading conclusions. The Critical Thinking feature provides additional clarification on specific concepts as a safeguard against incorrect evaluation of information.

Interpretation ▶

Increasingly, calculators and computers are used to generate the numeric results of a statistical process. However, the student still needs to correctly interpret those results in the context of a particular application. The Interpretation feature calls attention to this important step. Interpretation is stressed in examples, in guided exercises, and in the problem sets.

SOLUTION: Since we want to know the number of standard deviations from the mean, we want to convert 6.9 to standard z units.

$$z = \frac{x - \mu}{\sigma} = \frac{6.9 - 8}{0.5} = -2.20$$

Interpretation The amount of cheese on the selected pizza is only 2.20 standard deviations below the mean. The fact that z is negative indicates that the amount of cheese is 2.20 standard deviations *below* the mean. The parlor will not lose its franchise based on this sample.

6. *Interpretation* A campus performance series features plays, music groups, dance troops, and stand-up comedy. The committee responsible for selecting the performance groups include three students chosen at random from a pool of volunteers. This year the 30 volunteers came from a variety of majors. However, the three students for the committee were all music majors. Does this fact indicate there was bias in the selection process and that the selection process was not random? Explain.

7 *Critical Thinking* Greg took a random sample of size 100 from the population of current season ticket holders to State College men's basketball games. Then he took a random sample of size 100 from the population of current season ticket holders to State College women's basketball games.
   (a) What sampling technique (stratified, systematic, cluster, multistage, convenience, random) did Greg use to sample from the population of current season ticket holders to all State College basketball games played by either men or women?
   (b) Is it appropriate to pool the samples and claim to have a random sample of size 200 from the population of current season ticket holders to all State College home basketball games played by either men or women? Explain.

◀ *NEW!* Critical Thinking and Interpretation Exercises

In every section and chapter problem set, Critical Thinking problems provide students with the opportunity to test their understanding of the application of statistical methods and their interpretation of their results. Interpretation problems ask students to apply statistical results to the particular application.

STATISTICAL LITERACY

No language, including statistics, can be spoken without learning the vocabulary. *Understanding Basic Statistics* introduces statistical terms with deliberate care.

WHAT DOES THE LEVEL OF MEASUREMENT TELL US?

The level of measurement tells us which arithmetic processes are appropriate for the data. This is important because different statistical processes require various kinds of arithmetic. In some instances all we need to do is count the number of data that meet specified criteria. In such cases nominal (and higher) data levels are all appropriate. In other cases we need to order the data, so nominal data would not be suitable. Many other statistical processes require division, so data need to be at the ratio level. Just keep the nature of the data in mind before beginning statistical computations.

◀ *NEW!* What Does (concept, method, statistical result) Tell Us?

This feature gives a brief summary of the information we obtain from the named concept, method, or statistical result.

NEW! Important Features of a (concept, method, or result) ▶

In statistics we use many different types of graphs, samples, data, and analytical methods. The features of each such tool help us select the most appropriate ones to use and help us interpret the information we receive from applications of the tools.

IMPORTANT FEATURES OF A SIMPLE RANDOM SAMPLE

For a simple random sample

- Every sample of specified size *n* from the population has an equal chance of being selected.
- No researcher bias occurs in the items selected for the sample.
- A random sample may not always reflect the diversity of the population. For instance, from a population of 10 cats and 10 dogs, a random sample of size 6 could consist of all cats.

SECTION 7.1 PROBLEMS

1. *Statistical Literacy* Which, if any, of the curves in Figure 7-7 look(s) like a normal curve? If a curve is not a normal curve, tell why.

2. *Statistical Literacy* Look at the normal curve in Figure 7-8, and find μ, $\mu + \sigma$, and σ.

FIGURE 7-7

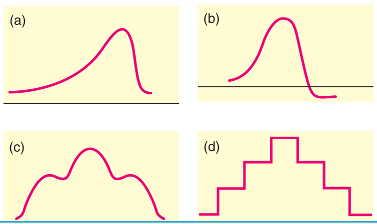
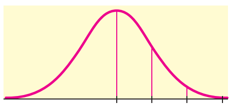


FIGURE 7-8



◀ Statistical Literacy Problems

In every section and chapter problem set, Statistical Literacy problems test student understanding of terminology, statistical methods, and the appropriate conditions for use of the different processes.

Definition Boxes ▶

Whenever important terms are introduced in text, tan definition boxes appear within the discussions. These boxes make it easy to reference or review terms as they are used further.

BOX-AND-WHISKER PLOTS

The quartiles together with the low and high data values give us a very useful *five-number summary* of the data and their spread.

FIVE-NUMBER SUMMARY

Lowest value, Q_1, median, Q_3, highest value

We will use these five numbers to create a graphic sketch of the data called a *box-and-whisker plot.* Box-and-whisker plots provide another useful technique from exploratory data analysis (EDA) for describing data.

STATISTICAL LITERACY

IMPORTANT WORDS & SYMBOLS

Section 5.1
Probability of an event A, $P(A)$ 190
Intuition 190
Relative frequency 190
Equally likely outcomes 190
Law of large numbers 192
Statistical experiment 192
Event 192
Simple event 192
Sample space 192
Complement of event A^c 194

Section 5.2
Independent events 202
Dependent events 202
Probability of A *and* B 202
Event $A \mid B$
Conditional probability 202
$P(A \mid B)$ 202
Multiplication rules of probability (for independent and dependent events) 202
More than two independent events 207
Probability of A *or* B 207
Event A *and* B 207
Event A *or* B 207
Mutually exclusive events 209
Addition rules (for mutually exclusive and general events) 209
More than two mutually exclusive events 211
Contingency table 211
Basic probability rules 214

Section 5.3
Multiplication rule of counting 221
Tree diagram 222
Factorial notation 225
Permutations rule 225
Combinations rule 227

◀ Important Words & Symbols

The Important Words & Symbols within the Chapter Review feature at the end of each chapter summarizes the terms introduced in the Definition Boxes for student review at a glance. Page numbers for first occurrence of term are given for easy reference.

Linking Concepts: Writing Projects ▶

Much of statistical literacy is the ability to communicate concepts effectively. The Linking Concepts: Writing Projects feature at the end of each chapter tests both statistical literacy and critical thinking by asking the student to express their understanding in words.

LINKING CONCEPTS: WRITING PROJECTS

Discuss each of the following topics in class or review the topics on your own. Then write a brief but complete essay in which you summarize the main points. Please include formulas and graphs as appropriate.

1. What does it mean to say that we are going to use a sample to draw an inference about a population? Why is a random sample so important for this process? If we wanted a random sample of students in the cafeteria, why couldn't we just choose the students who order Diet Pepsi with their lunch? Comment on the statement, "A random sample is like a miniature population, whereas samples that are not random are likely to be biased." Why would the students who order Diet Pepsi with lunch not be a random sample of students in the cafeteria?
2. In your own words, explain the differences among the following sampling techniques: simple random sample, stratified sample, systematic sample, cluster sample, multistage sample, and convenience sample. Describe situations in which each type might be useful.

5. *Basic Computation: Central Limit Theorem* Suppose x has a distribution with a mean of 8 and a standard deviation of 16. Random samples of size $n = 64$ are drawn.
   (a) Describe the $\bar{x}$ distribution and compute the mean and standard deviation of the distribution.
   (b) Find the z value corresponding to $\bar{x} = 9$.
   (c) Find $P(\bar{x} > 9)$.
   (d) *Interpretation* Would it be unusual for a random sample of size 64 from the x distribution to have a sample mean greater than 9? Explain.

◀ Basic Computation Problems

These problems focus student attention on relevant formulas, requirements, and computational procedures. After practicing these skills, students are more confident as they approach real-world applications.

Expand Your Knowledge Problems ▶

Expand Your Knowledge problems present optional enrichment topics that go beyond the material introduced in a section. Vocabulary and concepts needed to solve the problems are included at point-of-use, expanding students' statistical literacy.

30. *Expand Your Knowledge: Geometric Mean* When data consist of percentages, ratios, growth rates, or other rates of change, the *geometric mean* is a useful measure of central tendency. For n data values,

$$\text{Geometric mean} = \sqrt[n]{\text{product of the } n \text{ data values}}, \quad \text{assuming all data values are positive}$$

To find the *average growth factor* over 5 years of an investment in a mutual fund with growth rates of 10% the first year, 12% the second year, 14.8% the third year, 3.8% the fourth year, and 6% the fifth year, take the geometric mean of 1.10, 1.12, 1.148, 1.038, and 1.16. Find the average growth factor of this investment.

Note that for the same data, the relationships among the harmonic, geometric, and arithmetic means are harmonic mean $\le$ geometric mean $\le$ arithmetic mean (Source: *Oxford Dictionary of Statistics*).

DIRECTION AND PURPOSE

Real knowledge is delivered through direction, not just facts. *Understanding Basic Statistics* ensures the student knows what is being covered and why at every step along the way to statistical literacy.

Chapter Preview ▶ Questions

Preview Questions at the beginning of each chapter give the student a taste of what types of questions can be answered with an understanding of the knowledge to come.

NORMAL CURVES AND SAMPLING DISTRIBUTIONS

Dana White/PhotoEdit

PREVIEW QUESTIONS

What are some characteristics of a normal distribution? What does the empirical rule tell you about data spread around the mean? (SECTION 7.1)

Can you compare apples and oranges, or maybe elephants and butterflies? In most cases, the answer is no—unless you first standardize your measurements. What are a standard normal distribution and a standard *z* score? (SECTION 7.2)

How do you convert any normal distribution to a standard normal distribution? How do you find probabilities of "standardized events"? (SECTION 7.3)

As humans, our experiences are finite and limited. Consequently, most of the important decisions in our lives are based on sample (incomplete) information. What is a probability sampling distribution? How will sampling distributions help us make good decisions based on incomplete information? (SECTION 7.4)

There is an old saying: All roads lead to Rome. In statistics, we could recast this saying: All probability distributions average out to be normal distributions (as the sample size increases). How can we take advantage of this in our study of sampling distributions? (SECTION 7.5)

The binomial and normal distributions are two of the most important probability distributions in statistics. Under certain limiting conditions, the binomial can be thought to evolve (or envelope) into the normal distribution. How can you apply this concept in the real world? (SECTION 7.6)

Many issues in life come down to success or failure. In most cases, we will not be successful all the time, so proportions of successes are very important. What is

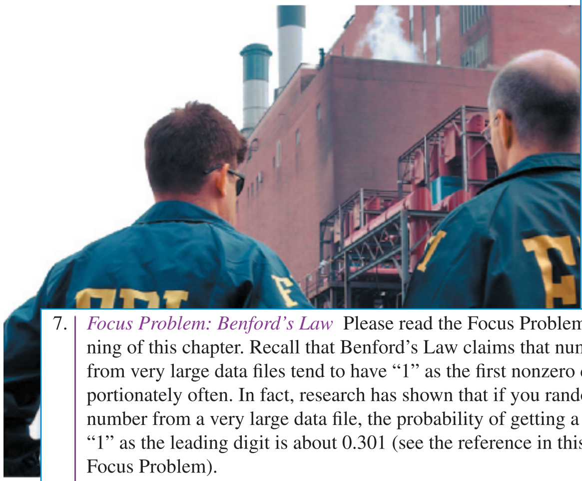
FOCUS PROBLEM

Benford's Law: The Importance of Being Number 1

Benford's Law states that in a wide variety of circumstances, numbers have "1" as their first nonzero digit disproportionately often. Benford's Law applies to such diverse topics as the drainage areas of rivers; properties of chemicals; populations of towns; figures in newspapers, magazines, and government reports; and the half-lives of radioactive atoms!

Specifically, such diverse measurements begin with "1" about 30% of the time, with "2" about 18% of time, and with "3" about 12.5% of the time. Larger digits occur less often. For example, less than 5% of the numbers in circumstances such as these begin with the digit 9. This is in dramatic contrast to a random sampling situation, in which each of the digits 1 through 9 has an equal chance of appearing.

The first nonzero digits of numbers taken from large bodies of numerical records such



Reuters/Corbis

7. *Focus Problem: Benford's Law* Please read the Focus Problem at the beginning of this chapter. Recall that Benford's Law claims that numbers chosen from very large data files tend to have "1" as the first nonzero digit disproportionately often. In fact, research has shown that if you randomly draw a number from a very large data file, the probability of getting a number with "1" as the leading digit is about 0.301 (see the reference in this chapter's Focus Problem).

   Now suppose you are an auditor for a very large corporation. The revenue report involves millions of numbers in a large computer file. Let us say you took a random sample of $n = 215$ numerical entries from the file and $r = 46$ of the entries had a first nonzero digit of 1. Let p represent the population proportion of all numbers in the corporate file that have a first nonzero digit of 1.

   i. Test the claim that p is less than 0.301. Use $\alpha = 0.01$.
   ii. If p is in fact less than 0.301, would it make you suspect that there are not enough numbers in the data file with leading 1's? Could this indicate that the books have been "cooked" by "pumping up" or inflating the numbers? Comment from the viewpoint of a stockholder. Comment from the perspective of the Federal Bureau of Investigation as it looks for money laundering in the form of false profits.

▲ Chapter Focus Problems

The Preview Questions in each chapter are followed by a Focus Problem, which serves as a more specific example of what questions the student will soon be able to answer. The Focus Problems are set within appropriate applications and are incorporated into the end-of-section exercises, giving students the opportunity to test their understanding.

DIRECTION AND PURPOSE

Focus Points ▶

Each section opens with bulleted Focus Points describing the primary learning objectives of the section.

SECTION **3.1**

Measures of Central Tendency: Mode, Median, and Mean

FOCUS POINTS

- Compute mean, median, and mode from raw data.
- Interpret what mean, median, and mode tell you.
- Explain how mean, median, and mode can be affected by extreme data values.
- What is a trimmed mean? How do you compute it?
- Compute a weighted average.

The average price of an ounce of gold is $1350. The Zippy car averages 39 miles per gallon on the highway. A survey showed the average shoe size for women is size 9.

In each of the preceding statements, *one* number is used to describe the entire sample or population. Such a number is called an *average*. There are many ways to compute averages, but we will study only three of the major ones.

Average

The easiest average to compute is the *mode*.

Mode

The **mode** of a data set is the value that occurs most frequently. *Note:* If a data set has no single value that occurs more frequently than any other, then that data set has no mode.

EXAMPLE 1

Mode

Count the letters in each word of this sentence and give the mode. The numbers of letters in the words of the sentence are

LOOKING FORWARD

In later chapters we will use information based on a sample and sample statistics to estimate population parameters (Chapter 8) or make decisions about the value of population parameters (Chapter 9).

◀ Looking Forward

This feature shows students where the presented material will be used later. It helps motivate students to pay a little extra attention to key topics.

CHAPTER REVIEW

SUMMARY

In this chapter, you've seen that statistics is the study of how to collect, organize, analyze, and interpret numerical information from populations or samples. This chapter discussed some of the features of data and ways to collect data. In particular, the chapter discussed

- Individuals or subjects of a study and the variables associated with those individuals
- Data classification as qualitative or quantitative, and levels of measurement of data
- Sample and population data. Summary measurements from sample data are called statistics, and those from populations are called parameters.
- Sampling strategies, including simple random, stratified, systematic, multistage, and convenience. Inferential techniques presented in this text are based on simple random samples.
- Methods of obtaining data: Use of a census, simulation, observational studies, experiments, and surveys
- Concerns: Undercoverage of a population, nonresponse, bias in data from surveys and other factors, effects of confounding or lurking variables on other variables, generalization of study results beyond the population of the study, and study sponsorship

▲ Chapter Summaries

The Summary within each Chapter Review feature now also appears in bulleted form, so students can see what they need to know at a glance.

REAL-WORLD SKILLS

Statistics is not done in a vacuum. *Understanding Basic Statistics* gives students valuable skills for the real world with technology instruction, genuine applications, actual data, and group projects.

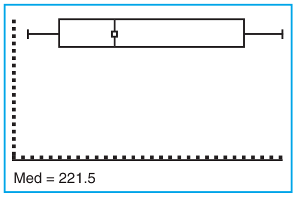
REVISED! Tech Notes ▶

Tech Notes appearing throughout the text give students helpful hints on using TI-84Plus and TI-*n*spire (with TI-84Plus keypad) and TI-83Plus calculators, Microsoft Excel 2010, and Minitab to solve a problem. They include display screens to help students visualize and better understand the solution.

TECH NOTES *Box-and-Whisker Plot*

Both Minitab and the TI-84Plus/TI-83Plus/TI-*n*spire calculators support box-and-whisker plots. On the TI-84Plus/TI-83Plus/TI-*n*spire, the quartiles Q_1 and Q_3 are calculated as we calculate them in this text. In Minitab and Excel 2010, they are calculated using a slightly different process.

TI-84Plus/TI-83Plus/TI-*n*spire (with TI-84Plus Keypad) Press STATPLOT ➤On. Highlight box plot. Use **Trace** and the arrow keys to display the values of the five-number summary. The display shows the plot for calories in ice cream bars.



box-and-whisker plot. However, each value of the
und. On the **Home** ribbon, click the **Insert Function**
tatistical as the category and scroll to **Quartile**. In
location and then enter the number of the value you
e quartile box for the first quartile.
t. In the dialogue box, set Display to IQRange Box.

USING TECHNOLOGY

Binomial Distributions

Although tables of binomial probabilities can be found in most libraries, such tables are often inadequate. Either the value of p (the probability of success on a trial) you are looking for is not in the table, or the value of n (the number of trials) you are looking for is too large for the table. In Chapter 6, we will study the normal approximation to the binomial. This approximation is a great help in many practical applications. Even so, we sometimes use the formula for the binomial probability distribution on a computer or graphing calculator to compute the probability we want.

Applications

The following percentages were obtained over many years of observation by the U.S. Weather Bureau. All data listed are for the month of December.

Location	Long-Term Mean % of Clear Days in Dec.
Juneau, Alaska	18%
Seattle, Washington	24%
Hilo, Hawaii	36%
Honolulu, Hawaii	60%
Las Vegas, Nevada	75%
Phoenix, Arizona	77%

Adapted from *Local Climatological Data*, U.S. Weather Bureau publication, "Normals, Means, and Extremes" Table.

In the locations listed, the month of December is a relatively stable month with respect to weather. Since weather patterns from one day to the next are more or less the same, it is reasonable to use a binomial probability model.

1. Let r be the number of clear days in December. Since December has 31 days, $0 \le r \le 31$. Using appropriate computer software or calculators available to you, find the probability $P(r)$ for each of the listed locations when $r = 0, 1, 2, \ldots, 31$.
2. For each location, what is the expected value of the probability distribution? What is the standard deviation?

You may find that using cumulative probabilities and appropriate subtraction of probabilities, rather than addition of probabilities, will make finding the solutions to Applications 3 to 7 easier.

3. Estimate the probability that Juneau will have at most 7 clear days in December.
4. Estimate the probability that Seattle will have from 5 to 10 (including 5 and 10) clear days in December.
5. Estimate the probability that Hilo will have at least 12 clear days in December.
6. Estimate the probability that Phoenix will have 20 or more clear days in December.
7. Estimate the probability that Las Vegas will have from 20 to 25 (including 20 and 25) clear days in December.

Technology Hints

TI-84Plus/TI-83Plus/TI-*n*spire (with TI-84 Plus keypad), Excel 2010, Minitab

The Tech Note in Section 5.2 gives specific instructions for binomial distribution functions on the TI-84Plus/TI-83Plus/TI-*n*spire (with TI-84Plus keypad) calculators, Excel 2010, and Minitab.

SPSS

In SPSS, the function **PDF.BINOM(q,n,p)** gives the probability of q successes out of n trials, where p is the probability of success on a single trial. In the data editor, name a variable r and enter values 0 through n. Name another variable Prob_r. Then use the menu choices **Transform ➤ Compute**. In the dialogue box, use Prob_r for the target variable. In the function group, select **PDF and Noncentral PDF**. In the function box, select **PDF.BINOM(q,n,p)**. Use the variable r for q and appropriate values for n and p. Note that the function **CDF.BINOM(q,n,p)**, from the **CDF and Noncentral CDF** group, gives the cumulative probability of 0 through q successes.

◀ *REVISED!* Using Technology

Further technology instruction is available at the end of each chapter in the Using Technology section. Problems are presented with real-world data from a variety of disciplines that can be solved by using TI-84Plus and TI-*n*spire (with TI-84Plus keypad) and TI-83Plus calculators, Microsoft Excel 2010, and Minitab.

REAL-WORLD SKILLS

EXAMPLE 11

Central limit theorem

A certain strain of bacteria occurs in all raw milk. Let x be the bacteria count per milliliter of milk. The health department has found that if the milk is not contaminated, then x has a distribution that is more or less mound-shaped and symmetrical. The mean of the x distribution is $\mu = 2500$, and the standard deviation is $\sigma = 300$. In a large commercial dairy, the health inspector takes 42 random samples of the milk produced each day. At the end of the day, the bacteria count in each of the 42 samples is averaged to obtain the sample mean bacteria count $\bar{x}$.

(a) Assuming the milk is not contaminated, what is the distribution of $\bar{x}$?

SOLUTION: The sample size is $n = 42$. Since this value exceeds 30, the central limit theorem applies, and we know
mean and standard deviation

◀ *UPDATED!* Applications

Real-world applications are used from the beginning to introduce each statistical process. Rather than just crunching numbers, students come to appreciate the value of statistics through relevant examples.

11. *Pain Management: Laser Therapy* "Effect of Helium-Neon Laser Auriculotherapy on Experimental Pain Threshold" is the title of an article in the journal *Physical Therapy* (Vol. 70, No. 1, pp. 24–30). In this article, laser therapy was discussed as a useful alternative to drugs in pain management of chronically ill patients. To measure pain threshold, a machine was used that delivered low-voltage direct current to different parts of the body (wrist, neck, and back). The machine measured current in milliamperes (mA). The pretreatment experimental group in the study had an average threshold of pain (pain was first detectable) at $\mu = 3.15$ mA with standard deviation $\sigma = 1.45$ mA. Assume that the distribution of threshold pain, measured in milliamperes, is symmetrical and more or less mound-shaped. Use the empirical rule to
    (a) estimate a range of milliamperes centered about the mean in which about 68% of the experimental group had a threshold of pain.
    (b) estimate a range of milliamperes centered about the mean in which about 95% of the experimental group had a threshold of pain.

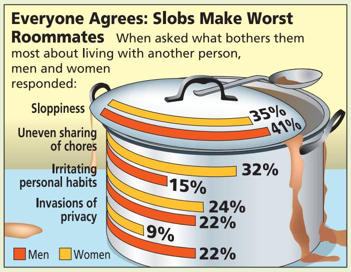
Most exercises in each section ▶ are applications problems.

DATA HIGHLIGHTS: GROUP PROJECTS

Break into small groups and discuss the following topics. Organize a brief outline in which you summarize the main points of your group discussion.

1. Examine Figure 2-20, "Everyone Agrees: Slobs Make Worst Roommates." This is a clustered bar graph because two percentages are given for each response category: responses from men and responses from women. Comment about how the artistic rendition has slightly changed the format of a bar graph. Do the bars seem to have lengths that accurately reflect the relative percentages of the responses? In your own opinion, does the artistic rendition enhance or confuse the information? Explain. Which characteristic of "worst roommates" does the graphic seem to illustrate? Can this graph be considered a Pareto chart for men? for women? Why or why not? From the information given in the figure, do you think the survey just listed the four given annoying characteristics? Do you think a respondent could choose more than one characteristic? Explain

FIGURE 2-20



Source: Advantage Business Research for Mattel *Compatibility*

◀ Data Highlights: Group Projects

Using Group Projects, students gain experience working with others by discussing a topic, analyzing data, and collaborating to formulate their response to the questions posed in the exercise.

MAKING THE JUMP

Get to the "Aha!" moment faster. *Understanding Basic Statistics* provides the push students need to get there through guidance and example.

PROCEDURE

HOW TO TEST μ WHEN σ IS KNOWN

Requirements

Let x be a random variable appropriate to your application. Obtain a simple random sample (of size n) of x values from which you compute the sample mean $\bar{x}$. The value of σ is already known (perhaps from a previous study). If you can assume that x has a normal distribution, then any sample size n will work. If you cannot assume this, then use a sample size $n \geq 30$.

Procedure

1. In the context of the application, state the *null and alternate hypotheses* and set the *level of significance* α.
2. Use the known σ, the sample size n, the value of $\bar{x}$ from the sample, and μ from the null hypothesis to compute the standardized *sample test statistic.*

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

3. Use the standard normal ... or two-tailed, to find the ...
4. *Conclude* the test. If *P*-va... do not reject H_0.
5. *Interpret your conclusion* ...

◀ Procedures and Requirements

Procedure display boxes summarize simple step-by-step strategies for carrying out statistical procedures and methods as they are introduced. Requirements for using the procedures are also stated. Students can refer to these boxes as they practice using the procedures.

GUIDED EXERCISE 10 **PROBABILITY REGARDING $\bar{x}$**

In mountain country, major highways sometimes use tunnels instead of long, winding roads over high passes. However, too many vehicles in a tunnel at the same time can cause a hazardous situation. Traffic engineers are studying a long tunnel in Colorado. If x represents the time for a vehicle to go through the tunnel, it is known that the x distribution has mean $\mu = 12.1$ minutes and standard deviation $\sigma = 3.8$ minutes under ordinary traffic conditions. From a histogram of x values, it was found that the x distribution is mound-shaped with some symmetry about the mean.

Engineers have calculated that, *on average,* vehicles should spend from 11 to 13 minutes in the tunnel. If the time is less than 11 minutes, traffic is moving too fast for safe travel in the tunnel. If the time is more than 13 minutes, there is a problem of bad air quality (too much carbon monoxide and other pollutants).

Under ordinary conditions, there are about 50 vehicles in the tunnel at one time. What is the probability that the mean time for 50 vehicles in the tunnel will be from 11 to 13 minutes?
We will answer this question in steps.

(a) Let $\bar{x}$ represent the sample mean based on samples of size 50. Describe the $\bar{x}$ distribution.

➡ From the central limit theorem, we expect the $\bar{x}$ distribution to be approximately normal, with mean and standard deviation

$$\mu_{\bar{x}} = \mu = 12.1 \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.8}{\sqrt{50}} \approx 0.54$$

(b) Find $P(11 < \bar{x} < 13)$.

Jupiterimages/Stockbyte/Getty Images

➡ We convert the interval

$11 < \bar{x} < 13$

to a standard z interval and use the standard normal probability table to find our answer. Since

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \approx \frac{\bar{x} - 12.1}{0.54}$$

$$\bar{x} = 11 \text{ converts to } z \approx \frac{11 - 12.1}{0.54} = -2.04$$

$$\text{and } \bar{x} = 13 \text{ converts to } z \approx \frac{13 - 12.1}{0.54} = 1.67$$

Therefore,

$$\begin{aligned} P(11 < \bar{x} < 13) &= P(-2.04 < z < 1.67) \\ &= 0.9525 - 0.0207 \\ &= 0.9318 \end{aligned}$$

(c) ***Interpret*** your answer to part (b).

➡ It seems that about 93% of the time, there should be no safety hazard for average traffic flow.

Guided Exercises ▶

Students gain experience with new procedures and methods through Guided Exercises. Beside each problem in a Guided Exercise, a completely worked-out solution appears for immediate reinforcement.

Preface

Welcome to the exciting world of statistics! We have written this text to make statistics accessible to everyone, including those with a limited mathematics background. Statistics affects all aspects of our lives. Whether we are testing new medical devices or determining what will entertain us, applications of statistics are so numerous that, in a sense, we are limited only by our own imagination in discovering new uses for statistics.

OVERVIEW

The seventh edition of *Understanding Basic Statistics* continues to emphasize *concepts of statistics*. Statistical methods are carefully presented with a focus on understanding both the *suitability of the method* and the *meaning of the result*. Statistical methods and measurements are developed in the context of applications.

Critical thinking and interpretation are essential in understanding and evaluating information. Statistical literacy is fundamental for applying and comprehending statistical results. In this edition we have expanded and highlighted the treatment of statistical literacy, critical thinking, and interpretation.

We have retained and expanded features that made the first six editions of the text very readable. Definition boxes highlight important terms. Procedure displays summarize steps for analyzing data. Examples, exercises, and problems touch on applications appropriate to a broad range of interests.

The seventh edition continues to have extensive online support. Online homework powered by a choice of Enhanced WebAssign or Aplia™ is now available through CengageBrain.com. Instructional videos are available on DVD. The companion web site at http://www.cengagebrain.com contains more than 100 data sets (in JMP, Microsoft Excel, Minitab, SPSS, and TI-84Plus/TI-83Plus/TI-*n*spire with TI-84Plus keypad ASCII file formats), technology guides, lecture aids, a glossary, and statistical tables.

MAJOR CHANGES IN THE SEVENTH EDITION

With each new edition, the authors reevaluate the scope, appropriateness, and effectiveness of the text's presentation and reflect on extensive user feedback. Revisions have been made throughout the text to clarify explanations of important concepts and to update problems.

JMP Statistical Software

All new copies of the text now come with JMP statistical software at no additional cost. We realize the vital importance of having a data analysis tool and understand the desirability for students to acquire familiarity with a commercial software package whose use might continue outside of this course. The text can still be used with many popular software packages as well. See Instructor Resources on page xxiii for a full description of the software.

New for the seventh edition, available via Aplia™, is MindTap™ Reader, Cengage Learning's next-generation eBook. MindTap Reader provides robust opportunities for students to annotate, take notes, navigate, and interact with the text (e.g., ReadSpeaker). Annotations captured in MindTap Reader are automatically tied to the Notepad app, where they can be viewed chronologically and in a cogent, linear fashion. Instructors also can edit the text and assets in the Reader as well as add videos or URLs. Go to http://www.cengage.com/mindtap for more information.

Critical Thinking, Interpretation, and Statistical Literacy

The seventh edition of this text continues and expands the emphasis on critical thinking, interpretation, and statistical literacy. Calculators and computers are very good at providing numerical results of statistical processes. However, numbers from a computer or calculator display are meaningless unless the user knows how to interpret the results and if the statistical process is appropriate. This text helps students determine whether or not a statistical method or process is appropriate. It helps students understand what a statistic measures. It helps students interpret the results of a confidence interval, hypothesis test, or liner regression model.

New Interpretation Features

To further understanding and interpretation of statistical concepts, methods, and results, we have included two new special features: **What Does (a concept, method, or result) Tell Us?** and **Important Features of a (concept, method, or result).** These features summarize the information we obtain from concepts and statistical processes and give additional insights for further application.

New Expand Your Knowledge Problems and Quick Overview Topics With Additional Applications

Expand Your Knowledge problems do just that! These are optional but contain very useful information taken from the vast literature of statistics. These topics are not included in the main text but are easily learned using material from the section or previous sections. Although these topics are optional, the authors feel they add depth and enrich a student's learning experience. Each topic was chosen for its relatively straightforward presentation and useful applications. All such problems and their applications are flagged with a sun logo.

New *Expand Your Knowledge problems* in the seventh edition provide additional topics found in Advanced Placement courses. These topics include logarithmic transformation problems featuring exponential growth and power law applications, linear functions of random variables, and confidence intervals for the slope of the least-squares line. Also new are problems involving stratified sampling and the best estimate for the population mean μ.

Some of the other topics in *Expand Your Knowledge problems* or quick overviews include graphs such as dotplots, donut graphs, and variations on stem-and-leaf plots; outliers in stem-and-leaf plots; harmonic and geometric means; moving averages; calculating odds in favor and odds against; continuous uniform distribution, quick estimate of the standard deviation using the Empirical rule; plus four confidence intervals for proportions; Satterthwaite's approximation for degrees of freedom in confidence intervals and hypothesis tests; relationship between confidence intervals and two-tailed hypothesis testing; pooled two-sample procedures for confidence intervals and hypothesis tests; resampling (also known as bootstrap); simulations of confidence intervals and hypothesis tests using different samples of the same size; and serial correlation (also called autocorrelation).

For location of these optional topics in the text, please see the index.

Revised Examples and New Section Problems

Examples and guided exercises have been updated and revised. Additional section problems emphasize critical thinking and interpretation of statistical results.

Excel 2010 and Most Recent Operating System for the TI-84Plus/TI-83Plus Calculators

Excel 2010 instructions are included in the *Tech Notes* and *Using Technology*. The latest operating system (v2.55MP) for the TI-84Plus/TI-83Plus calculators is also discussed, with new functions such as the inverse t distribution and the chi-square goodness of fit test described. One convenient feature of the operating system is that it provides on-screen prompts for inputs required for many probability and statistical functions. This operating system is already on new TI-84Plus/TI-83Plus calculators and is available for download to older calculators at the Texas Instruments web site.

Revised Electronic Student Resources

Digital student resources and online tools that accompany *Understanding Basic Statistics* have been revised in accordance with recommendations from both student and faculty users. New copies of the text come with JMP statistical software. Online interactive learning solutions, such as Aplia™ for Statistics—featuring the new MindTap™ Reader—and Enhanced WebAssign, are both available.

CONTINUING CONTENT

Introduction of Hypothesis Testing Using *P*-Values

In keeping with the use of computer technology and standard practice in research, hypothesis testing is introduced using P-values. The critical region method is still supported but not given primary emphasis.

Use of Student's *t* Distribution in Confidence Intervals and Testing of Means

If the normal distribution is used in confidence intervals and testing of means, then the *population standard deviation must be known.* If the population standard deviation is not known, then under conditions described in the text, the Student's t distribution is used. This is the most commonly used procedure in statistical research. It is also used in statistical software packages such as JMP, Microsoft Excel, Minitab, SPSS, and TI-84Plus/TI-83Plus/TI-*n*spire calculators.

Confidence Intervals and Hypothesis Tests of Difference of Means

If the normal distribution is used, then both population standard deviations must be known. When this is not the case, the Student's t distribution incorporates an approximation for t, with a commonly used conservative choice for the degrees of freedom. Satterthwaite's approximation for the degrees of freedom as used in computer software is also discussed. The pooled standard deviation is presented for appropriate applications ($\sigma_1 \approx \sigma_2$).

FEATURES IN THE SEVENTH EDITION

Chapter and Section Lead-ins

- *Preview Questions* at the beginning of each chapter are keyed to the sections.
- *Focus Problems* at the beginning of each chapter demonstrate types of questions students can answer once they master the concepts and skills presented in the chapter.
- *Focus Points* at the beginning of each section describe the primary learning objectives of the section.

Carefully Developed Pedagogy

- *Examples* show students how to select and use appropriate procedures.
- *Guided Exercises* within the sections give students an opportunity to work with a new concept. Completely worked-out solutions appear beside each exercise to give immediate reinforcement.
- *Definition boxes* highlight important definitions throughout the text.
- *Procedure displays* summarize key strategies for carrying out statistical procedures and methods. Conditions required for using the procedure are also stated.
- NEW! *What Does (a concept method or result) Tell Us?* summarizes information we obtain from the named concepts and statistical processes and gives insight for additional application.
- NEW! *Important Features of a (concept, method, or result)* summarizes the features of the listed item.
- *Looking Forward* features give a brief preview of how a current topic is used later.
- *Labels* for each example or guided exercise highlight the technique, concept, or process illustrated by the example or guided exercise. In addition, labels for section and chapter problems describe the field of application and show the wide variety of subjects in which statistics is used.
- *Section and chapter problems* require the student to use all the new concepts mastered in the section or chapter. Problem sets include a variety of real-world applications with data or settings from identifiable sources. Key steps and solutions to odd-numbered problems appear at the end of the book.
- *Basic Computation problems* ask students to practice using formulas and statistical methods on very small data sets. Such practice helps students understand what a statistic measures.
- *Statistical Literacy problems* ask students to focus on correct terminology and processes of appropriate statistical methods. Such problems occur in every section and chapter problem set.
- *Interpretation problems* ask students to explain the meaning of the statistical results in the context of the application.
- *Critical Thinking problems* ask students to analyze and comment on various issues that arise in the application of statistical methods and in the interpretation of results. These problems occur in every section and chapter problem set.
- *Expand Your Knowledge problems* present enrichment topics such as dot plots, grouped data, estimation of standard deviation from a range of data values, residual plots, relationship between confidence intervals and two-tailed hypothesis tests, and more.
- *Cumulative review problem sets* occur after every third chapter and include key topics from previous chapters. Answers to *all* cumulative review problems are given at the end of the book.
- *Data Highlights and Linking Concepts* provide group projects and writing projects.
- *Viewpoints* are brief essays presenting diverse situations in which statistics is used.
- *Design and photos* are appealing and enhance readability.

Technology Within the Text

- *Tech Notes* within sections provide brief point-of-use instructions for the TI-84Plus, TI-83Plus, and TI-*n*spire (with 84Plus keypad) calculators, Microsoft Excel 2010, and Minitab.
- *Using Technology* sections show the use of SPSS as well as the TI-84Plus, TI-83Plus, and TI-*n*spire (with TI-84Plus keypad) calculators, Microsoft Excel, and Minitab.

ALTERNATE ROUTES THROUGH THE TEXT

Understanding Basic Statistics, Seventh Edition, is designed to be flexible. It offers the professor a choice of teaching possibilities. In most one-semester courses, it is not practical to cover all the material in depth. However, depending on the emphasis of the course, the professor may choose to cover various topics. For help in topic selection, refer to the Table of Prerequisite Material on page 1.

- *Linear regression.* Chapter 4, Correlation and Regression, may be delayed until after Chapter 9. The descriptive topics of linear regression may then be followed immediately by the inferential topics of linear regression presented in Chapter 11.
- *Probability.* For courses requiring minimal probability, Section 5.1 (What Is Probability?) and the first part of Section 5.2 (Some Probability Rules—Compound Events) will be sufficient.

ACKNOWLEDGMENTS

It is our pleasure to acknowledge the prepublication reviewers of this text. All of their insights and comments have been very valuable to us. Reviewers of this text include:

Kimberly Benien, Wharton County Junior College
Abraham Biggs, Broward Community College
Melissa Bingham, University of Wisconsin-La Crosse
Chand Chauhan, Indiana University-Purdue University-Fort Wayne
Tammy Hoyle, Gardner-Webb University
David Read, Paradise Valley Community College
Diane Van Deusen, Napa Valley College
Daniel Weiner, Boston University

We also would like to acknowledge reviewers of previous editions.

Reza Abbasian, Texas Lutheran University
Paul Ache, Kutztown University
Kathleen Almy, Rock Valley College
Polly Amstutz, University of Nebraska at Kearney
Delores Anderson, Truett-McConnell College
Robert J. Astalos, Feather River College
Lynda L. Ballou, Kansas State University
Mary Benson, Pensacola Junior College
Larry Bernett, Benedictine University
Kiran Bhutani, The Catholic University of America
Kristy E. Bland, Valdosta State University
John Bray, Broward Community College
Bill Burgin, Gaston College
Toni Carroll, Siena Heights University
Pinyuen Chen, Syracuse University
Emmanuel des-Bordes, James A. Rhodes State College
Jennifer M. Dollar, Grand Rapids Community College
Larry E. Dunham, Wor-Wic Community College
Andrew Ellett, Indiana University
Ruby Evans, Keiser University
Mary Fine, Moberly Area Community College
Rebecca Fouguette, Santa Rosa Junior College
Rene Garcia, Miami-Dade Community College
Larry Green, Lake Tahoe Community College
Shari Harris, John Wood Community College
Janice Hector, DeAnza College
Jane Keller, Metropolitan Community College
Raja Khoury, Collin County Community College
Diane Koenig, Rock Valley College
Charles G. Laws, Cleveland State Community College

Michael R. Lloyd, Henderson State University
Beth Long, Pellissippi State Technical and Community College
Lewis Lum, University of Portland
Darcy P. Mays, Virginia Commonwealth University
Charles C. Okeke, College of Southern Nevada, Las Vegas
Peg Pankowski, Community College of Allegheny County
Ram Polepeddi, Westwood College, Denver North Campus
Azar Raiszadeh, Chattanooga State Technical Community College
Traei Reed, St. Johns River Community College
Michael L. Russo, Suffolk County Community College
Janel Schultz, Saint Mary's University of Minnesota
Sankara Sethuraman, Augusta State University
Stephen Soltys, West Chester University of Pennsylvania
Ron Spicer, Colorado Technical University
Winson Taam, Oakland University
Jennifer L. Taggart, Rockford College
William Truman, University of North Carolina at Pembroke
Bill White, University of South Carolina Upstate
Jim Wienckowski, State University of New York at Buffalo
Stephen M. Wilkerson, Susquehanna University
Hongkai Zhang, East Central University
Shunpu Zhang, University of Alaska, Fairbanks
Cathy Zucco-Teveloff, Trinity College

We would especially like to thank Roger Lipsett for his careful accuracy review of this text. We are especially appreciative of the excellent work by the editorial and production professionals at Cengage Learning. In particular we thank Molly Taylor, Jay Campbell, and Jill Quinn.

Without their creative insight and attention to detail, a project of this quality and magnitude would not be possible. Finally, we acknowledge the cooperation of Minitab, Inc., SPSS, Texas Instruments, and Microsoft.

Charles Henry Brase

Corrinne Pellillo Brase

Additional Resources — Get More from Your Textbook!

JMP is a statistics software for Windows and Macintosh computers from SAS, the market leader in analytics software and services for industry. JMP Student Edition is a streamlined, easy-to-use version that provides all the statistical analysis and graphics covered in this textbook. Once data is imported, students will find that most procedures require just two or three mouse clicks. JMP can import data from a variety of formats, including Excel and other statistical packages, and you can easily copy and paste graphs and output into documents.

JMP also provides an interface to explore data visually and interactively, which will help your students develop a healthy relationship with their data, work more efficiently with data, and tackle difficult statistical problems more easily. Because its output provides both statistics and graphs together, the student will better see and understand the application of concepts covered in this book as well. JMP Student Edition also contains some unique platforms for student projects, such as mapping and scripting. JMP functions in the same way on both Windows and Macintosh platforms and instructions contained with this book apply to both platforms.

Access to this software is available with new copies of the book. Students can purchase JMP standalone via CengageBrain.com or www.jmp.com/getse.

Aplia™ for Statistics makes it easier for you to develop students' technical skills, conceptual understanding, and ability to interpret data, draw conclusions, and put the results in context. This balance makes it easier to teach students to think statistically, helping them appreciate the value of interpreting data in the world around them. Aplia™ is an online interactive learning solution that improves comprehension and outcomes by increasing student effort and engagement. Founded by a professor to enhance his own courses, Aplia™ provides automatically graded assignments with detailed, immediate explanations on every question, and innovative teaching materials. Our easy to use system has been used by more than 1,000,000 students at over 1,800 institutions. Exercises are taken directly from text. Aplia homework engages students in critical thinking, requiring them to synthesize and apply knowledge, not simply recall it. The diverse types of questions reflect the types of exercises that help students learn. All homework is written by subject matter experts in the field who have taught the course before. Aplia contains a robust course management system with powerful analytics, enabling professors to track student performance easily.

MindTap™

Available via Aplia™ is MindTap™ Reader, Cengage Learning's next-generation eBook. An integral component of the MindTap environment, MindTap Reader replaces the traditional "e-book" and provides robust opportunities for students to annotate, take notes, navigate, and interact with the text (e.g., ReadSpeaker). Annotations captured in MindTap Reader are automatically tied to the Notepad app, where they can be viewed chronologically and in a cogent, linear fashion. Instructors also can edit the text and assets in the Reader, as well as add videos or URLs.

Go to http://www.cengage.com/mindtap for more information.

Instructor Resources

Annotated Instructor's Edition (AIE) Answers to all exercises, teaching comments, and pedagogical suggestions appear in the margin, or at the end of the text in the case of large graphs.

Complete Solutions Manual This manual contains solutions to all exercises from the text, including Chapter Review Exercises, and Cumulative Review Exercises. This manual can be found on the Instructors Companion Site.

Cengage Learning Testing Powered by Cognero A flexible, online system that allows you to:

- author, edit, and manage test bank content from multiple Cengage Learning solutions
- create multiple test versions in an instant
- deliver tests from your LMS, your classroom or wherever you want

Companion Website The companion website at http://www.cengage.com/login contains a variety of resources.

- Microsoft® PowerPoint® lecture slides
- Figures from the book
- More than 100 data sets in a variety of formats, including
  - JMP
  - Microsoft Excel
  - Minitab
  - SPSS
  - TI-84Plus/TI-83Plus/TI-nspire with 84plus keypad ASCII file formats
- Technology guides for the following programs
  - JMP
  - TI-84Plus, TI-83Plus, and TI-nspire graphing calculators
  - Minitab software (version 14)
  - Microsoft Excel (2010/2007)
  - SPSS Statistics software
- Lecture aids like Teaching Hints and Frequently Used Formulas
- Statistical tables and a glossary

Student Resources

Student Solutions Manual Provides solutions to the odd-numbered section and chapter exercises and to all the Cumulative Review exercises in the student textbook.

Instructional DVDs Hosted by Dana Mosely, these text-specific DVDs cover all sections of the text and provide explanations of key concepts, examples, exercises, and applications in a lecture-based format. DVDs are close-captioned for the hearing-impaired.

Minitab® and IBM SPSS These statistical software packages manipulate and interpret data to produce textual, graphical, and tabular results. Minitab® and/or SPSS may be packaged with the textbook. Student versions are available.

The companion website at http://www.cengage.com/statistics/brase11e contains useful assets for students.

- **Technology Guides** Separate guides exist with information and examples for various technology tools. Guides are available for JMP; the TI-84Plus, TI-83Plus, and TI-*n*spire graphing calculators; Minitab software (version 14) Microsoft Excel (2010/2007); and SPSS Statistics software.
- **Interactive Teaching and Learning Tools** include glossary flashcards, online datasets (in JMP, Microsoft Excel, Minitab, SPSS, and Tl-84Plus/TI-83Plus/TI-*n*spire with TI-84Plus keypad ASCII file formats), statistical tables and formulas, and more.

Enhanced WebAssign Offers an extensive online program for Statistics to encourage the practice that's so critical for concept mastery. The skillfully crafted pedagogy and exercises in Brase and Brase's text become even more effective in Enhanced WebAssign.

CengageBrain.com Provides the freedom to purchase online homework and other materials à la carte—exactly what you need, when you need it.

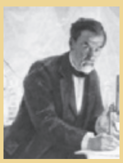
For more information, visit **http://www.cengage.com/statistics/brase7e** or contact your local Cengage Learning sales representative.

Table of Prerequisite Material

Chapter		Prerequisite Sections
1	Getting Started	None
2	Organizing Data	1.1, 1.2
3	Averages and Variation	1.1, 1.2, 2.1
4	Correlation and Regression	1.1, 1.2, 3.1, 3.2
5	Elementary Probability Theory	1.1, 1.2, 2.1
6	The Binomial Probability Distribution and Related Topics	1.1, 1.2, 2.1, 3.1, 3.2, 5.1, 5.2 5.3 useful but not essential
7	Normal Curves and Sampling	
	Distributions (omit 7.6)	1.1, 1.2, 2.1, 3.1, 3.2, 5.1, 5.2, 6.1
	(include 7.6)	also 6.2, 6.3
8	Estimation	
	(omit 8.3)	1.1, 1.2, 2.1, 3.1, 3.2, 5.1, 5.2, 6.1, 7.1, 7.2, 7.3, 7.4, 7.5
	(include 8.3)	also 6.2, 6.3, 7.6
9	Hypothesis Testing	
	(omit 9.3)	1.1, 1.2, 2.1, 3.1, 3.2, 5.1, 5.2, 6.1, 7.1, 7.2, 7.3, 7.4, 7.5
	(include 9.3)	also 6.2, 6.3, 7.6
10	Inferences about Differences	
	(omit 10.3)	1.1, 1.2, 2.1, 3.1, 3.2, 5.1, 5.2, 6.1, 7.1, 7.2, 7.3, 7.4, 7.5, 8.1, 8.2, 9.1, 9.2
	(include 10.3)	also 6.2, 6.3, 7.6, 9.3
11	Additional Topics Using Inference	
	(Part I: 11.1, 11.2, 11.3)	1.1, 1.2, 2.1, 3.1, 3.2, 5.1, 5.2, 6.1, 7.1, 7.2, 7.3, 7.4, 7.5, 9.1
	(Part II: 11.4)	Chapter 4, 8.1, 8.2, 9.1, 9.2 also

1

Paul Spinelli/Contributor/Major League Baseball/Getty Images



Bettmann/Historical/Corbis

Chance favors the prepared mind.

—Louis Pasteur

Statistical techniques are tools of thought . . . not substitutes for thought.

—Abraham Kaplan

Louis Pasteur (1822–1895) is the founder of modern bacteriology. At age 57, Pasteur was studying cholera. He accidentally left some bacillus culture unattended in his laboratory during the summer. In the fall, he injected laboratory animals with this bacilli. To his surprise, the animals did not die—in fact, they thrived and were resistant to cholera.

When the final results were examined, it is said that Pasteur remained silent for a minute and then exclaimed, as if he had seen a vision, "Don't you see they have been vaccinated!" Pasteur's work ultimately saved many human lives.

Most of the important decisions in life involve incomplete information. Such decisions often involve so many complicated factors that a complete analysis is not practical or even possible. We are often forced into the position of making a guess based on limited information.

As the first quote reminds us, our chances of success are greatly improved if we have a "prepared mind." The statistical methods you will learn in this book will help you achieve a prepared mind for the study of many different fields. The second quote reminds us that statistics is an important tool, but it is not a replacement for an in-depth knowledge of the field to which it is being applied.

The authors of this book want you to understand and enjoy statistics. The reading material will *tell you* about the subject. The examples will *show you* how it works. To understand, however, you must *get involved*. Guided exercises, calculator and computer applications, section and chapter problems, and writing exercises are all designed to get you involved in the subject. As you grow in your understanding of statistics, we believe you will enjoy learning a subject that has a world full of interesting applications.

GETTING STARTED

PREVIEW QUESTIONS

Why is statistics important? (SECTION 1.1)

What is the nature of data? (SECTION 1.1)

How can you draw a random sample? (SECTION 1.2)

What are other sampling techniques? (SECTION 1.2)

How can you design ways to collect data? (SECTION 1.3)

National News/ZUMA Press/Newscom

FOCUS PROBLEM

Where Have All the Fireflies Gone?

A feature article in *The Wall Street Journal* discusses the disappearance of fireflies. In the article, Professor Sara Lewis of Tufts University and other scholars express concern about the decline in the worldwide population of fireflies.

There are a number of possible explanations for the decline, including habitat reduction of woodlands, wetlands, and open fields; pesticides; and pollution. Artificial nighttime lighting might interfere with the Morse-code-like mating ritual of the fireflies. Some chemical companies pay a bounty for fireflies because the insects contain two rare chemicals used in medical research and electronic detection systems used in spacecraft.

Courtesy of Corrinne and Charles Brase

Adapted from Ohio State University Firefly Files logo

What does any of this have to do with statistics?

The truth, at this time, is that no one really knows (a) how much the world firefly population has declined or (b) how to explain the decline. The population of all fireflies is simply too large to study in its entirety.

In any study of fireflies, we must rely on incomplete information from samples. Furthermore, from these samples we must draw realistic conclusions that have statistical integrity. This is the kind of work that makes use of statistical methods to determine ways to collect, analyze, and investigate data.

Suppose you are conducting a study to compare firefly populations exposed to normal daylight/darkness conditions with firefly populations exposed to continuous light (24 hours a day). You set up two firefly colonies in a laboratory environment. The two colonies are identical except that one colony is exposed to normal daylight/darkness conditions and the other is exposed to continuous light. Each colony is populated with the same number of mature fireflies. After 72 hours, you count the number of living fireflies in each colony.

After completing this chapter, you will be able to answer the following questions.

(a) Is this an experiment or an observation study? Explain.

(b) Is there a control group? Is there a treatment group?

For online student resources, visit the Brase/Brase, *Understanding Basic Statistics*, 7th edition web site at http://www.cengagebrain.com

(c) What is the variable in this study?
(d) What is the level of measurement (nominal, interval, ordinal, or ratio) of the variable? (See Problem 11 of the Chapter 1 Review Problems.)

SECTION **1.1**

What Is Statistics?

FOCUS POINTS

- Identify variables in a statistical study.
- Distinguish between quantitative and qualitative variables.
- Identify populations and samples.
- Distinguish between parameters and statistics.
- Determine the level of measurement.
- Compare descriptive and inferential statistics.

INTRODUCTION

Decision making is an important aspect of our lives. We make decisions based on the information we have, our attitudes, and our values. Statistical methods help us examine information. Moreover, statistics can be used for making decisions when we are faced with uncertainties. For instance, if we wish to estimate the proportion of people who will have a severe reaction to a flu shot without giving the shot to everyone who wants it, statistics provides appropriate methods. Statistical methods enable us to look at information from a small collection of people or items and make inferences about a larger collection of people or items.

Procedures for analyzing data, together with rules of inference, are central topics in the study of statistics.

Statistics

Statistics is the study of how to collect, organize, analyze, and interpret numerical information from data.

The subject of statistics is multifaceted. The following definition of statistics is found in the *International Encyclopedia of Statistical Science*, edited by Miodrag Lovric. Professor David Hand of Imperial College London—the president of the Royal Statistical Society—presents the definition in his article "Statistics: An Overview."

Statistics is both the science of uncertainty and the technology of extracting information from data.

The statistical procedures you will learn in this book should supplement your built-in system of inference—that is, the results from statistical procedures and good sense should dovetail. Of course, statistical methods themselves have no power to work miracles. These methods can help us make some decisions, but not all conceivable decisions. Remember, even a properly applied statistical procedure is no more accurate than the data, or facts, on which it is based. Finally, statistical results should be interpreted by one who understands not only the methods, but also the subject matter to which they have been applied.

The general prerequisite for statistical decision making is the gathering of data. First, we need to identify the individuals or objects to be included in the study and the characteristics or features of the individuals that are of interest.

Individuals
Variable

Individuals are the people or objects included in the study.
A **variable** is a characteristic of the individual to be measured or observed.

For instance, if we want to do a study about the people who have climbed Mt. Everest, then the individuals in the study are all people who have actually made it to the summit. One variable might be the height of such individuals. Other variables might be age, weight, gender, nationality, income, and so on. Regardless of the variables we use, we would not include measurements or observations from people who have not climbed the mountain.

The variables in a study may be *quantitative* or *qualitative* in nature.

Quantitative variable
Qualitative variable

A **quantitative variable** has a value or numerical measurement for which operations such as addition or averaging make sense. A **qualitative variable** describes an individual by placing the individual into a category or group, such as male or female.

For the Mt. Everest climbers, variables such as height, weight, age, or income are *quantitative* variables. *Qualitative variables* involve nonnumerical observations such as gender or nationality. Sometimes qualitative variables are referred to as *categorical variables.*

Categorical variable

Another important issue regarding data is their source. Do the data comprise information from *all* individuals of interest, or from just *some* of the individuals?

Population data
Sample data

In **population data,** the data are from *every* individual of interest.
In **sample data,** the data are from *only some* of the individuals of interest.

It is important to know whether the data are population data or sample data. Data from a specific population are fixed and complete. Data from a sample may vary from sample to sample and are *not* complete.

Population parameter

Sample statistic

A **population parameter** is a numerical measure that describes an aspect of a population.
A **sample statistic** is a numerical measure that describes an aspect of a sample.

For instance, if we have data from *all* the individuals who have climbed Mt. Everest, then we have population data. The proportion of males in the *population* of all climbers who have conquered Mt. Everest is an example of a *parameter.*

On the other hand, if our data come from just some of the climbers, we have sample data. The proportion of male climbers in the *sample* is an example of a *statistic*. Note that different samples may have different values for the proportion of male climbers. One of the important features of sample statistics is that they can vary from sample to sample, whereas population parameters are fixed for a given population.

LOOKING FORWARD

In later chapters we will use information based on a sample and sample statistics to estimate population parameters (Chapter 8) or make decisions about the value of population parameters (Chapters 9 and 10).

EXAMPLE 1

Using basic terminology

The Hawaii Department of Tropical Agriculture is conducting a study of ready-to-harvest pineapples in an experimental field.

(a) The pineapples are the *objects* (individuals) of the study. If the researchers are interested in the individual weights of pineapples in the field, then the *variable* consists of weights. At this point, it is important to specify units of measurement

Gallo Images/Landbouweekblad/Getty Images

and degrees of accuracy of measurement. The weights could be measured to the nearest ounce or gram. Weight is a *quantitative* variable because it is a numerical measure. If weights of *all* the ready-to-harvest pineapples in the field are included in the data, then we have a *population*. The average weight of all ready-to-harvest pineapples in the field is a *parameter.*

(b) Suppose the researchers also want data on taste. A panel of tasters rates the pineapples according to the categories "poor," "acceptable," and "good." Only some of the pineapples are included in the taste test. In this case, the *variable* is taste. This is a *qualitative* or *categorical* variable. Because only some of the pineapples in the field are included in the study, we have a *sample*. The proportion of pineapples in the sample with a taste rating of "good" is a *statistic*.

Throughout this text, you will encounter *guided exercises* embedded in the reading material. These exercises are included to give you an opportunity to work immediately with new ideas. The questions guide you through appropriate analysis. Cover the answers on the right side (an index card will fit this purpose). After you have thought about or written down *your own response,* check the answers. If there are several parts to an exercise, check each part before you continue. You should be able to answer most of these exercise questions, but don't skip them—they are important.

GUIDED EXERCISE 1 Using basic terminology

Television station QUE wants to know the proportion of TV owners in Virginia who watch the station's new program at least once a week. The station asks a group of 1000 TV owners in Virginia if they watch the program at least once a week.

(a) Identify the individuals of the study and the variable.

→ The individuals are the 1000 TV owners surveyed. The variable is the response "does" or "does not" watch the new program at least once a week.

(b) Do the data comprise a sample? If so, what is the underlying population?

→ The data comprise a sample of the population of responses from all TV owners in Virginia.

(c) Is the variable qualitative or quantitative?

→ Qualitative—the categories are the two possible responses, "does" or "does not" watch the program.

(d) Identify a quantitative variable that might be of interest.

→ Age or income might be of interest.

(e) Is the proportion of viewers in the sample who watch the new program at least once a week a statistic or a parameter?

→ Statistic—the proportion is computed from sample data.

LEVELS OF MEASUREMENT: NOMINAL, ORDINAL, INTERVAL, RATIO

We have categorized data as either qualitative or quantitative. Another way to classify data is according to one of the four *levels of measurement.* These levels indicate the type of arithmetic that is appropriate for the data, such as ordering, taking differences, or taking ratios.

Levels of measurement

LEVELS OF MEASUREMENT

Nominal level

The **nominal level of measurement** applies to data that consist of names, labels, or categories. There are no implied criteria by which the data can be ordered from smallest to largest.

Ordinal level

The **ordinal level of measurement** applies to data that can be arranged in order. However, differences between data values either cannot be determined or are meaningless.

Interval level

The **interval level of measurement** applies to data that can be arranged in order. In addition, differences between data values are meaningful.

Ratio level

The **ratio level of measurement** applies to data that can be arranged in order. In addition, both differences between data values and ratios of data values are meaningful. Data at the ratio level have a true zero.

EXAMPLE 2

Levels of measurement

Identify the type of data.

(a) Taos, Acoma, Zuni, and Cochiti are the names of four Native American pueblos from the population of names of all Native American pueblos in Arizona and New Mexico.

SOLUTION: These data are at the *nominal* level. Notice that these data values are simply names. By looking at the name alone, we cannot determine if one name is "greater than or less than" another. Any ordering of the names would be numerically meaningless.

© Michelle Dulieu/Shutterstock.com

(b) In a high school graduating class of 319 students, Jim ranked 25th, June ranked 19th, Walter ranked 10th, and Julia ranked 4th, where 1 is the highest rank.

SOLUTION: These data are at the *ordinal* level. Ordering the data clearly makes sense. Walter ranked higher than June. Jim had the lowest rank, and Julia the highest. However, numerical differences in ranks do not have meaning. The difference between June's and Jim's ranks is 6, and this is the same difference that exists between Walter's and Julia's ranks. However, this difference doesn't really mean anything significant. For instance, if you looked at grade point average, Walter and Julia may have had a large gap between their grade point averages, whereas June and Jim may have had closer grade point averages. In any ranking system, it is only the relative standing that matters. Computed differences between ranks are meaningless.

© iStockphoto.com/Korban Schwabi

(c) Body temperatures (in degrees Celsius) of trout in the Yellowstone River.

SOLUTION: These data are at the *interval* level. We can certainly order the data, and we can compute meaningful differences. However, for Celsius-scale temperatures, there is not an inherent starting point. The value 0°C may seem to be a starting point, but this value does not indicate the state of "no heat." Furthermore, it is not correct to say that 20°C is twice as hot as 10°C.

(d) Length of trout swimming in the Yellowstone River.

SOLUTION: These data are at the *ratio* level. An 18-inch trout is three times as long as a 6-inch trout. Observe that we can divide 6 into 18 to determine a meaningful *ratio* of trout lengths.

In summary, there are four levels of measurement. The nominal level is considered the lowest, and in ascending order we have the ordinal, interval, and ratio levels. In general, calculations based on a particular level of measurement may not be appropriate for a lower level.

PROCEDURE

HOW TO DETERMINE THE LEVEL OF MEASUREMENT

The levels of measurement, listed from lowest to highest, are nominal, ordinal, interval, and ratio. To determine the level of measurement of data, state the *highest level* that can be justified for the entire collection of data. Consider which calculations are suitable for the data.

Level of Measurement	Suitable Calculation
Nominal	We can put the data into categories.
Ordinal	We can order the data from smallest to largest or "worst" to "best." Each data value can be *compared* with another data value.
Interval	We can order the data and also take the differences between data values. At this level, it makes sense to compare the differences between data values. For instance, we can say that one data value is 5 more than or 12 less than another data value.
Ratio	We can order the data, take differences, and also find the ratio between data values. For instance, it makes sense to say that one data value is twice as large as another.

WHAT DOES THE LEVEL OF MEASUREMENT TELL US?

The level of measurement tells us which arithmetic processes are appropriate for the data. This is important because different statistical processes require various kinds of arithmetic. In some instances all we need to do is count the number of data that meet specified criteria. In such cases nominal (and higher) data levels are all appropriate. In other cases we need to order the data, so nominal data would not be suitable. Many other statistical processes require division, so data need to be at the ratio level. Just keep the nature of the data in mind before beginning statistical computations.

GUIDED EXERCISE 2 LEVELS OF MEASUREMENT

The following describe different data associated with a state senator. For each data entry, indicate the corresponding *level of measurement.*

(a) The senator's name is Sam Wilson. → Nominal level

(b) The senator is 58 years old. → Ratio level. Notice that age has a meaningful zero. It makes sense to give age ratios. For instance, Sam is twice as old as someone who is 29.

Continued

GUIDED EXERCISE 2 *continued*

(c) The years in which the senator was elected to the Senate are 2000, 2006, and 2012.	Interval level. Dates can be ordered, and the difference between dates has meaning. For instance, 2006 is 6 years later than 2000. However, ratios do not make sense. The year 2000 is not twice as large as the year 1000. In addition, the year 0 does not mean "no time."
(d) The senator's total taxable income last year was \$878,314.	Ratio level. It makes sense to say that the senator's income is 10 times that of someone earning \$87,831.40.
(e) The senator surveyed his constituents regarding his proposed water protection bill. The choices for response were strong support, support, neutral, against, or strongly against.	Ordinal level. The choices can be ordered, but there is no meaningful numerical difference between two choices.
(f) The senator's marital status is "married."	Nominal level
(g) A leading news magazine claims the senator is ranked seventh for his voting record on bills regarding public education.	Ordinal level. Ranks can be ordered, but differences between ranks may vary in meaning.

CRITICAL THINKING

"Data! Data! Data!" he cried impatiently. "I can't make bricks without clay." Sherlock Holmes said these words in *The Adventure of the Copper Beeches* by Sir Arthur Conan Doyle.

Reliable statistical conclusions require reliable data. This section has provided some of the vocabulary used in discussing data. As you read a statistical study or conduct one, pay attention to the nature of the data and the ways they were collected.

When you select a variable to measure, be sure to specify the process and requirements for measurement. For example, if the variable is the weight of ready-to-harvest pineapples, specify the unit of weight, the accuracy of measurement, and maybe even the particular scale to be used. If some weights are in ounces and others in grams, the data are fairly useless.

Another concern is whether or not your measurement instrument truly measures the variable. Just asking people if they know the geographic location of the island nation of Fiji may not provide accurate results. The answers may reflect the fact that the respondents want you to think they are knowledgeable. Asking people to locate Fiji on a map may give more reliable results.

The level of measurement is also an issue. You can put numbers into a calculator or computer and do all kinds of arithmetic. However, you need to judge whether the operations are meaningful. For ordinal data such as restaurant rankings, you can't conclude that a 4-star restaurant is "twice as good" as a 2-star restaurant, even though the number 4 is twice 2.

Continued