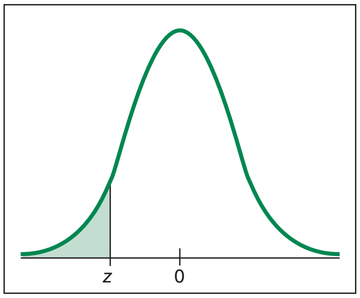
Cengage

13th Edition

Understandable Statistics

Concepts and Methods

BRASE • BRASE
DOLOR • SEIBERT



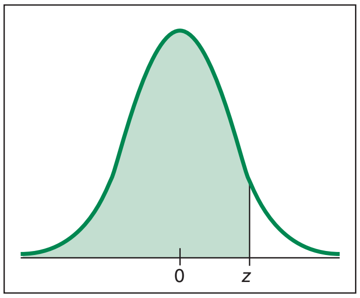
The table entry for *z* is the area to the left of *z*.

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.



The 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

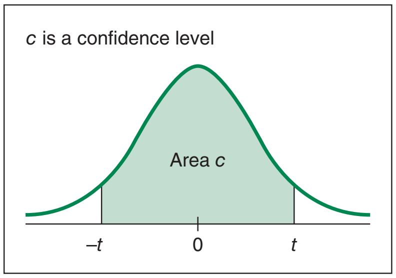
For z values greater than 3.49, use 1.000 to approximate the area.

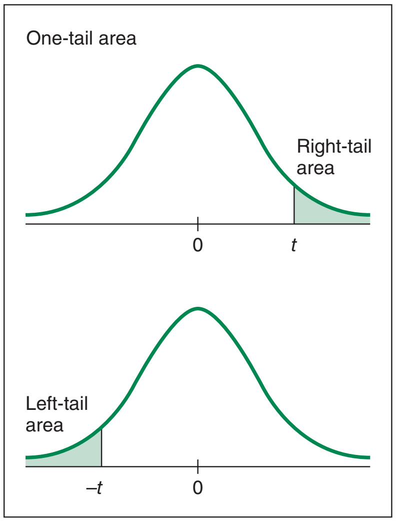
Areas of a Standard Normal Distribution *continued*

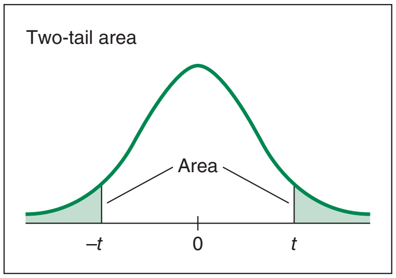
(b) Confidence Interval Critical Values z_c	
Level of Confidence c	**Critical Value z_c**
0.70, or 70%	1.04
0.75, or 75%	1.15
0.80, or 80%	1.28
0.85, or 85%	1.44
0.90, or 90%	1.645
0.95, or 95%	1.96
0.98, or 98%	2.33
0.99, or 99%	2.58

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







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	1.178	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*.

FREQUENTLY USED FORMULAS

f = frequency	s = sample standard deviation	$\hat{p}$ = sample proportion
n = sample size	σ = population standard deviation	p = population proportion
N = population size	s^2 = sample variance	p = probability of success
$\bar{x}$ = sample mean	σ^2 = population variance	q = probability of failure
μ = population mean		

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

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 5

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

Standard deviation of a discrete probability distribution
$\sigma = \sqrt{\Sigma(x-\mu)^2 P(x)}$

Given $L = a + bx$
$\mu_L = a + b\mu$
$\sigma_L = |b|\sigma$

Given $W = ax_1 + bx_2$ (x_1 and x_2 independent)
$\mu_W = a\mu_1 + b\mu_2$
$\sigma_W = \sqrt{a^2\sigma_1^2 + b^2\sigma_2^2}$

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 $\sigma = \sqrt{npq}$

Geometric Probability Distribution

n = number of trial on which first success occurs
$P(n) = p(1-p)^{n-1}$

Poisson Probability Distribution

r = number of successes

λ = mean number of successes over given interval

$P(r) = \dfrac{e^{-\lambda}\lambda^r}{r!}$

Chapter 6

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 of $\hat{p}$ distribution $\sigma_{\hat{p}} = \sqrt{\dfrac{pq}{n}}$; $q = 1 - p$

Chapter 7

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

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

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

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

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

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 8

Sample Test Statistics for Tests of Hypotheses

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

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

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

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

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

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 $\bar{q} = 1 - \bar{p}$

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

Chapter 9

Regression and Correlation

Pearson product-moment correlation coefficient

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

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

where $b = \dfrac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$; also $b = r\left(\dfrac{s_x}{s_y}\right)$

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

Coefficient of determination $= r^2$

Sample test statistic for r

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

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

Confidence interval for y

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

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

with $d.f. = n - 2$

Sample test statistic for slope b

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

Confidence interval for β

$$b - E < \beta < b + E$$

$$\text{where } E = \frac{t_c S_e}{\sqrt{\sum x^2 - \frac{1}{n}(\sum x)^2}} \text{ with } d.f. = n - 2$$

Chapter 10

$$\chi^2 = \sum \frac{(O - E)^2}{E} \text{ where}$$

O = observed frequency and
E = expected frequency

For tests of independence and tests of homogeneity

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

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) $- 1$

Confidence interval for σ^2; $d.f. = n - 1$

$$\frac{(n-1)s^2}{\chi^2_U} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}$$

Sample test statistic for σ^2

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

Testing Two Variances

Sample test statistic $F = \frac{s_1^2}{s_2^2}$

where $s_1^2 \geq s_2^2$

$d.f._N = n_1 - 1$; $d.f._D = n_2 - 1$

ANOVA

k = number of groups; N = total sample size

$$SS_{TOT} = \sum x_{TOT}^2 - \frac{(\sum x_{TOT})^2}{N}$$

$$SS_{BET} = \sum_{\text{all groups}} \left(\frac{(\sum x_i)^2}{n_i}\right) - \frac{(\sum x_{TOT})^2}{N}$$

$$SS_W = \sum_{\text{all groups}} \left(\sum x_i^2 - \frac{(\sum x_i)^2}{n_i}\right)$$

$$SS_{TOT} = SS_{BET} + SS_W$$

$$MS_{BET} = \frac{SS_{BET}}{d.f._{BET}} \text{ where } d.f._{BET} = k - 1$$

$$MS_W = \frac{SS_W}{d.f._W} \text{ where } d.f._W = N - k$$

$$F = \frac{MS_{BET}}{MS_W} \text{ where } d.f. \text{ numerator} = d.f._{BET} = k - 1;$$

$$d.f. \text{ denominator} = d.f._W = N - k$$

Two-Way ANOVA

r = number of rows; c = number of columns

Row factor F: $\frac{MS \text{ row factor}}{MS \text{ error}}$

Column factor F: $\frac{MS \text{ column factor}}{MS \text{ error}}$

Interaction F: $\frac{MS \text{ interaction}}{MS \text{ error}}$

with degrees of freedom for

row factor $= r - 1$ interaction $= (r - 1)(c - 1)$

column factor $= c - 1$ error $= rc(n - 1)$

Chapter 11

Sample test statistic for x = proportion of plus signs to all signs ($n \geq 12$)

$$z = \frac{x - 0.5}{\sqrt{0.25/n}}$$

Sample test statistic for R = sum of ranks

$$z = \frac{R - \mu_R}{\sigma_R} \text{ where } \mu_R = \frac{n_1(n_1 + n_2 + 1)}{2} \text{ and}$$

$$\sigma_R = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}$$

Spearman rank correlation coefficient

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)} \text{ where } d = x - y$$

Sample test statistic for runs test

R = number of runs in sequence

THIRTEENTH EDITION

Understandable Statistics

CONCEPTS AND METHODS

Charles Henry Brase

REGIS UNIVERSITY

Corrinne Pellillo Brase

ARAPAHOE COMMUNITY COLLEGE

Jason Dolor

PORTLAND STATE UNIVERSITY

James Seibert

REGIS UNIVERSITY

Cengage

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

***Understandable Statistics: Concepts and Methods*, Thirteenth Edition**
Charles Henry Brase, Corrinne Pellillo Brase, Jason Dolor, and James Seibert

SVP, Higher Education Product Management: Erin Joyner

VP, Product Management, Learning Experiences: Thais Alencar

Product Director: Mark Santee

Product Managers: Andy Trus, Spencer Arritt

Product Assistant: Kyra Kruger

Senior Learning Designer: Elinor Gregory

Content Managers: Abigail DeVeuve, Amanda Rose, Adrian Daniel

Digital Delivery Quality Partner: Nikkita Kendrick

Senior Director, Product Marketing: Jennifer Fink

Product Marketing Manager: Shaunna Campbell

IP Analyst: Ashley Maynard

IP Project Manager: Carly Belcher

Production Service: Straive

Designers: Nadine Ballard, Gaby Vinales

Cover Image Source: gettyimages.com/Byrdyak

For product information and technology assistance, contact us at
Cengage Customer & Sales Support, 1-800-354-9706
or support.cengage.com.

For permission to use material from this text or product,
submit all requests online at **www.copyright.com.**

Library of Congress Control Number: 2021921577

ISBN: 978-0-357-71917-6

Cengage
200 Pier 4 Boulevard
Boston, MA 02210
USA

Cengage is a leading provider of customized learning solutions with employees residing in nearly 40 different countries and sales in more than 125 countries around the world. Find your local representative at **www.cengage.com.**

To learn more about Cengage platforms and services, register or access your online learning solution, or purchase materials for your course, visit **www.cengage.com.**

Notice to the Reader

Publisher does not warrant or guarantee any of the products described herein or perform any independent analysis in connection with any of the product information contained herein. Publisher does not assume, and expressly disclaims, any obligation to obtain and include information other than that provided to it by the manufacturer. The reader is expressly warned to consider and adopt all safety precautions that might be indicated by the activities described herein and to avoid all potential hazards. By following the instructions contained herein, the reader willingly assumes all risks in connection with such instructions. The publisher makes no representations or warranties of any kind, including but not limited to, the warranties of fitness for particular purpose or merchantability, nor are any such representations implied with respect to the material set forth herein, and the publisher takes no responsibility with respect to such material. The publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or part, from the readers' use of, or reliance upon, this material.

Printed in the United States of America
Print Number: 02 Print Year: 2022

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

Burton W. Jones
Professor Emeritus, University of Colorado

Contents

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 thirteenth edition of *Understandable Statistics: Concepts and Methods* continues to emphasize concepts of statistics that are covered in Introductory Statistics courses. 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. Updated Critical Thinking activities give students opportunities to deeply explore concepts through hands-on learning that challenge student thinking beyond procedural fluency. Revised Viewpoint boxes also help students synthesize what they've learned by engaging with real data and applying concepts to real-world situations.

We have retained and expanded features that made the first 12 editions of the text very readable. Definition boxes highlight important terms. Procedure displays summarize steps for analyzing data. Examples, exercises, and problems have been updated for currency, relevancy, and an increased focus on diversity, equity, and inclusion. Additionally, the Cengage Instructor Center at faculty.cengage.com contains more than 100 data sets and technology guides.

WEBASSIGN

WebAssign for Brase/Brase's *Understandable Statistics: Concepts and Methods*, Thirteenth Edition, puts powerful tools in the hands of instructors with a flexible and fully customizable online instructional solution, enabling them to deploy assignments, instantly assess individual student and class performance, and help students master the course concepts. With WebAssign's powerful digital platform and *Understandable Statistics*' specific content, instructors can tailor their course with a wide range of assignment settings, add their own questions and content, and connect with students effectively using communication tools.

Major Changes in the Thirteenth 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, engage students in discussion and active learning using simulations, and help all students feel included in the content.

Global Updates

- Contexts throughout the text have been updated to improve diversity, equity, and inclusion.
- Examples, Guided Exercises, and Problems have been updated for currency and relevancy.
- Critical Thinking and Viewpoint boxes have been revised to engage students in discussions and hands-on learning using simulations and real data, including data from Cengage's Dataset Hub.
- SALT (Statistical Analysis and Learning Tool) has been incorporated into the Tech Notes and Using Technology sections.
- Over 100 new exercises have been added.
- Expand Your Knowledge has been streamlined to incorporate relevant content into the main text and remove content that was beyond the scope of the learning objectives.

Chapter Updates

- Chapter 2: Organizing Data
  - New Focus Problem on Covid-19.
- Chapter 3: Averages and Variation
  - New introduction to Standard Deviation in Section 3.2.
  - New coverage on Grouped Data in Section 3.2.
- Chapter 4: Elementary Probability Theory
  - New explanation of the Law of Large Numbers in Section 4.1.
  - New discussion of events with very high or very low probabilities in Section 4.1.
  - New Critical Thinking simulation activity on probability in Section 4.1.
- Chapter 5: The Binomial Probability Distribution and Related Topics
  - New Focus Problem on Experiencing Other Cultures.
  - Removed coverage of "Linear Functions of a Random Variable" and "Linear Combinations of Independent Random Variables" in Section 5.1.
  - Removed coverage of "Sampling Without Replacement: Use of the Hypergeometric Probability Distribution" in Section 5.2.
  - Section 5.4 updated and available in the etextbook only.
- Chapter 6: Normal Curves and Sampling Distributions
  - Introduction to sampling distributions significantly expanded to include a $\hat{p}$ distribution example and exercises assessing the normality distributions in Section 6.4.
  - New Critical Thinking activity on $\hat{p}$ distribution in Section 6.4.
  - New Critical Thinking simulation activity on sampling distributions in Section 6.4.
- Chapter 7: Estimation
  - New Critical Thinking simulation activity on confidence intervals in Section 7.1.
  - New Critical Thinking simulation activity on how confidence level, sample size, and sample proportion impact a confidence interval for a proportion in Section 7.3.
- Chapter 8: Hypothesis Testing
  - Expanded explanation of p-value with accompanying Critical Thinking activity in Section 8.1.
  - New Critical Thinking simulation activity on Hypothesis Testing in Section 8.3.
- Chapter 9: Correlation and Regression
  - New Critical Thinking simulation activity on the effects of outliers on regression and correlation in Section 9.2.

Continuing Content

Critical Thinking, Interpretation, and Statistical Literacy

The thirteenth 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 linear regression model.

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 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 Thirteenth 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.
- *Learning Objectives* at the beginning of each section describe what students should be able to do after completing 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 terminology throughout the text.
- *Procedure displays* summarize key strategies for carrying out statistical procedures and methods. Conditions required for using the procedure are also stated.
- *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.
- *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.
- *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* present real data in context and ask students to analyze and interpret the data using what they've learned.
- *Critical Thinking* activities strengthen conceptual understanding by engaging students in discussions and hands-on learning using simulations.

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, SALT, 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, SALT, and Minitab.

Alternate Routes Through the Text

Understandable Statistics: Concepts and Methods, Thirteenth 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.

- *Introducing linear regression early.* For courses requiring an early presentation of linear regression, the descriptive components of linear regression (Sections 9.1 and 9.2) can be presented any time after Chapter 3. However, inference topics involving predictions, the correlation coefficient ρ, and the slope of the least-squares line β require an introduction to confidence intervals (Sections 7.1 and 7.2) and hypothesis testing (Sections 8.1 and 8.2).
- *Probability.* For courses requiring minimal probability, Section 4.1 (What Is Probability?) and the first part of Section 4.2 (Some Probability Rules—Compound Events) will be sufficient.

Instructor Resources

Additional resources for this product are available in the Cengage Instructor Center. Instructor assets include a Complete Solutions Manual, PowerPoint® slides, Guide to Teaching Online, Educator's Guide, a test bank powered by Cognero®, and more. Sign up or sign in at faculty.cengage.com to search for and access this product and its online resources.

- Complete Solutions Manual—provides solutions and answers to textbook questions.
- PowerPoint® slides—support lectures with definitions, formulas, examples, and activities.
- Guide to Teaching Online—offers tips for teaching online and incorporating WebAssign activities into your course.
- Educator's Guide—offers suggested content for WebAssign by chapter to help you personalize your course.
- Cengage Testing, powered by Cognero®—a flexible, online system that allows you to access, customize, and deliver a test bank from your chosen text to your students through your LMS or another channel outside of Webassign.
- Data Sets—provide the data from textbook exercises in downloadable files.
- Technology Guides—help students work through problems using TI-83, TI-84, TI-*N*spire calculators, Excel, Minitab, and SPSS.
- Transition Guide—outlines changes between the 12th and 13th editions of the textbook.

Acknowledgments

It is our pleasure to acknowledge all of the reviewers, past and present, who have helped make this book what it is over its thirteen 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
Peter Arvanites, Rockland Community College
Robert J. Astalos, Feather River College
Jorge Baca, Cosumnes River College
Lynda L. Ballou, Kansas State University
Wayne Barber, Chemeketa Community College
Molly Beauchman, Yavapai College
Nick Belloit, Florida State College at Jacksonville
Kimberly Benien, Wharton County Junior College
Mary Benson, Pensacola Junior College
Larry Bernett, Benedictine University
Kiran Bhutani, The Catholic University of America
Abraham Biggs, Broward Community College
Kristy E. Bland, Valdosta State University
John Bray, Broward Community College
Bill Burgin, Gaston College
Dexter Cahoy, Louisiana Tech University
Maggy Carney, Burlington County 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
Christopher Donnelly, Macomb Community College

Kathleen Donoghue, Robert Morris University
Larry E. Dunham, Wor-Wic Community College
Andrew Ellett, Indiana University
Ruby Evans, Keiser University
Sherrie Fenner, Lehigh Carbon Community College
Mary Fine, Moberly Area Community College
Rebecca Fouguette, Santa Rosa Junior College
Rene Garcia, Miami-Dade Community College
Nicholas Gorgievski, Nichols College
Larry Green, Lake Tahoe Community College
Shari Harris, John Wood Community College
Janice Hector, DeAnza College
Tammy Hoyle, Gardner-Webb University
Jane Keller, Metropolitan Community College
Raja Khoury, Collin County Community College
Diane Koenig, Rock Valley College
Jenna Kowalski, Anoka-Ramsey Community College
Kewal Krishan, Hudson County Community College
Charles G. Laws, Cleveland State Community College
Tracy Leshan, Baltimore City 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
Cindy Moss, San Mateo County Community College
Meike Niederhausen, University of Portland
Charles C. Okeke, College of Southern Nevada, Las Vegas
Peg Pankowski, Community College of Allegheny County
Deanna Payton, Northern Oklahoma College in Stillwater
Michael Perry, Loyola University Chicago
Ram Polepeddi, Westwood College, Denver North Campus
Azar Raiszadeh, Chattanooga State Technical Community College
Jennifer Reeb, Oakton Community College
Traei Reed, St. Johns River Community College
Michael L. Russo, Suffolk County Community College
Alan Safer, California State University, Long Beach
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
Michelle Van Wagoner, Nashville State Community College
Bill White, University of South Carolina Upstate
Jim Wienckowski, State University of New York at Buffalo
Stephen M. Wilkerson, Susquehanna University
Ping Ye, University of North Georgia
Hongkai Zhang, East Central University
Shunpu Zhang, University of Alaska, Fairbanks
Cathy Zucco-Teveloff, Trinity College

The authors of the current edition, Jason Dolor and James Seibert, would especially like to thank Elinor Gregory for her creative insight and attention to detail. We are indebted to the entire Cengage Learning team that helped us through this revision. In particular we would like to acknowledge Andy Trus, Spencer Arritt, Abby DeVeuve, Amanda Rose, and Adrian Daniel. Special thanks are due to Charles and Corrinne Brase, who trusted us with this revision of their book!

About the Authors

Charles Brase had more than 30 years of full-time teaching experience in mathematics and statistics. He taught at the University of Hawaii, Manoa Campus, for several years and at Regis University in Denver, Colorado, for more than 28 years. Charles received the Excellence in Teaching award from the University of Hawaii and the Faculty Member of the Year award from Regis University on two occasions. He earned degrees from the University of Colorado, Boulder, and has a Ph.D. in Mathematics, an M.A. in Mathematics, and a B.A. in Physics. Charles worked on a combined 20 editions of *Understandable Statistics* and *Understanding Basic Statistics* before his passing in 2017.

Corrinne Pellillo Brase has taught at Hawaii Pacific College, Honolulu Community College, and Arapahoe Community College in Littleton, Colorado. She was also involved in the mathematics component of an equal opportunity program at the University of Colorado. Corrinne received the Faculty of the Year award from Arapahoe Community College. She earned degrees from the University of Colorado, Boulder, and has an M.A. and B.A. in Mathematics.

Jason Dolor has more than 16 years of teaching experience in Mathematics and Statistics. He briefly taught at the University of Guam and is currently a teacher and researcher at Portland State University and University of Portland, where he has been for the last 15 years. Jason has publications in research journals and has presented at education conferences on his work in statistical thinking, teacher knowledge, educational technology, curriculum development, and assessment. Jason earned his B.S. in Computer Science from the University of Portland, Graduate Certificate in Applied Statistics, M.S. in Mathematics, and a Ph.D. in Mathematics Education from Portland State University.

James Seibert has more than 20 years of full-time teaching experience in Mathematics and Statistics. He taught briefly at Colorado State University and Willamette University, and currently teaches at Regis University in Denver, Colorado, where he has been for the last 21 years. James was mentored early in his career at Regis by Charles Brase, and remained friends with both Charles and Corrinne Brase ever since. James earned his B.A. in Mathematics with minors in Physics and Philosophy from Linfield College, and his M.A. and Ph.D. in Mathematics from Colorado State University.

Critical Thinking

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

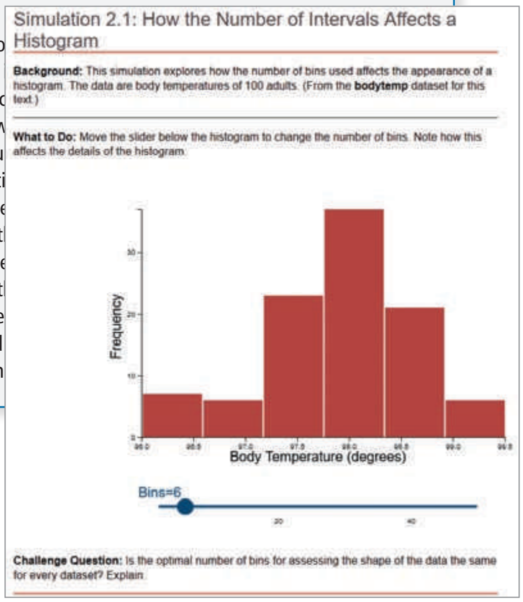
CRITICAL THINKING

As researchers, it is important to consider the number of classes when generating a histogram since it will help determine how the data is presented. When producing a histogram, it is considered good practice that the graph is presented in such a way that the reader is able to get a clear understanding of the data. Using a small number of classes may cause information to be hidden within particular bars. Too many classes may result in a histogram that could potentially overwhelm a reader. This activity will have you understand how to determine the appropriate number of classes for a specific data set.

Use Simulation 2.1: How the Number of Intervals Affects a Histogram to answer the questions below. (The simulation is also available on the Resources tab in WebAssign).

You see a histogram sho
adults. Below the histogram
ber of classes (i.e., bins) whic
slider, generate histograms w
the graphs change. After you
consider the following questi

1. What did you notice
ing a histogram wit
2. What did you notice
ing a histogram wit
3. If you were to prese
do you think would
Explain your reason



◀ UPDATED! Critical Thinking Boxes

Critical thinking is an important skill for students to develop in order to avoid reaching misleading conclusions. The Critical Thinking feature gives students the opportunity to explore specific concepts by engaging them in activities and discussion to help them internalize the topics and prepare them to anticipate issues that might arise.

Interpretation ▶

Students need to correctly interpret results in the context of a particular application. The Interpretation feature calls attention to this important step. Interpretation is stressed in examples, guided exercises, and problem sets.

5. **Interpretation** Suppose you are a hospital manager and have been told that there is no need to worry that respirator monitoring equipment might fail because the probability any one monitor will fail is only 0.01. The hospital has 20 such monitors and they work independently. Should you be more concerned about the probability that *exactly one* of the 20 monitors fails, or that *at least one* fails? Explain.

11. **Critical Thinking** Consider two data sets with equal sample standard deviations. The first data set has 20 data values that are not all equal, and the second has 50 data values that are not all equal. For which data set is the difference between s and σ greater? Explain. *Hint:* Consider the relationship $\sigma = s\sqrt{(n-1)/n}$.

◀ Critical Thinking Problems

Critical Thinking problems provide students with the opportunity to test their understanding of the application of statistical methods and their interpretation of their results.

Statistical Literacy

No language, including statistics, can be spoken without learning the vocabulary. *Understandable Statistics: Concepts and Methods* introduces statistical terms with deliberate care.

What Do Counting Rules Tell Us?

Counting rules tell us the total number of outcomes created by combining a sequence of events in specified ways.

- The **multiplication rule of counting** tells us the total number of possible outcomes for a sequence of events.
- **Tree diagrams** provide a visual display of all the resulting outcomes.
- The **permutation rule** tells us the total number of ways we can **arrange in order** n distinct objects into a group of size r.
- The **combination rule** tells us how many ways we can form n distinct objects into a group of size r, where the order is irrelevant.

◀ 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.

4. **Statistical Literacy** Let A = the event someone tested positive for a virus and B = the event someone has the virus. What is the contextual difference between $P(A \mid B)$ and $P(A \textit{ and } B)$?

◀ Statistical Literacy Problems

Statistical Literacy problems test student understanding of terminology, statistical methods, and the appropriate conditions for use of the different processes.

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. In your own words, explain the differences among histograms, relative-frequency histograms, bar graphs, circle graphs, time-series graphs, Pareto charts, and stem-and-leaf displays. If you have nominal data, which graphic displays might be useful? What if you have ordinal, interval, or ratio data?
2. What do we mean when we say a histogram is skewed to the left? To the right? What is a bimodal histogram? Discuss the following statement: "A bimodal histogram usually results if we draw a sample from two populations at once." Suppose you took a sample of weights of college football players and with this sample you included weights of cheerleaders. Do you think a histogram made from the combined weights would be bimodal? Explain.
3. Discuss the statement that stem-and-leaf displays are quick and easy to construct. How can we use a stem-and-leaf display to make the construction of a frequency table easier? How does a stem-and-leaf display help you spot extreme values quickly?
4. Go to the library and pick up a current issue of *The Wall Street Journal, Newsweek, Time, USA Today*, or other news medium. Examine each newspaper or magazine for graphs of the types discussed in this chapter. List the variables used, method of data collection, and general types of conclusions drawn from the graphs. Another source for information is the Internet. Explore several web sites, and categorize the graphs you find as you did for the print media.

9. **Basic Computation: Range, Standard Deviation**
Consider the data set

11 12 13 20 30

(a) Find the range.
(b) Use the defining formula to compute the sample standard deviation s.
(c) Use the defining formula to compute the population standard deviation σ.

◀ 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. More Basic Computation problems have been added to most sections.

Direction and Purpose

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

UPDATED! 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.

Standret/Shutterstock.com

PREVIEW QUESTIONS

How can we use probability to analyze events in life that are uncertain? (SECTION 4.1)
What are the basic definitions and rules of probability? (SECTION 4.2)
What are counting techniques, trees, permutations, and combinations? (SECTION 4.3)

FOCUS PROBLEM

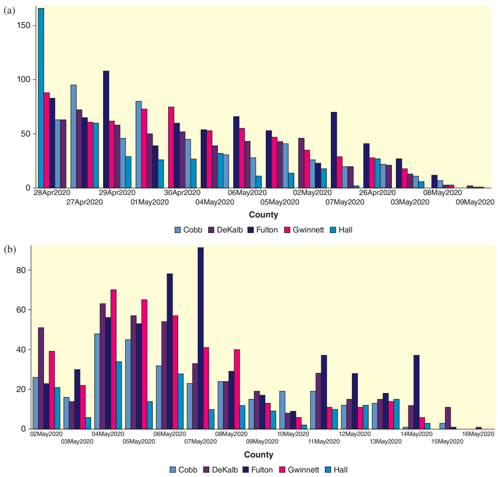
Say It with Pictures

Edward R. Tufte, in his book *The Visual Display of Quantitative Information*, presents a number of guidelines for producing good graphics. According to the criteria, a graphical display should

- show the data;
- induce the viewer to think about the substance of the graphic rather than the methodology, the design, the technology, or other production devices;
- avoid distorting what the data have to say.

Tufte includes a graphic that appeared in a well-known newspaper in his book that violates some of the criteria. Figure 2-1(a) shows another problematic graph that was originally posted on the *Georgia Department of Public Health* web site in 2020 during the height of the COVID-19 pandemic. The graphic was meant to represent the number of confirmed COVID-19 cases in the top five counties in Georgia over a 15-day span. Several critics and news sites suggested that the graph was misleading. As a result, the graph was quickly removed and replaced with a corrected graph, Figure 2-1(b).

FIGURE 2-1



◀ Chapter Focus Problems

Focus Problems introduce chapter concepts in real-world contexts and highlight questions students will be able to answer after completing the chapter. Focus Problems are incorporated into the end of section exercises, giving students an opportunity to test their understanding.

Learning Objectives ▶

Learning Objectives identify what students should be able to do after completing each section.

SECTION 4.1 **What Is Probability?**

LEARNING OBJECTIVES

- **Assign probabilities to events in the real world.**
- **Explain how the law of large numbers relates to relative frequencies.**
- **Apply basic rules of probability in everyday life.**
- **Explain the relationship between statistics and probability.**

We encounter statements given in terms of probability all the time. An excited sports announcer claims that Sheila has a 90% chance of breaking the world record in the upcoming 100-yard dash. Henry figures that if he guesses on a true–false question, the probability of getting it right is 1/2. A pharmaceutical company claimed that their new flu vaccine developed by their scientists has an efficacy rate of 0.91.

When we use probability in a statement, we're using a *number between 0 and 1* to indicate the likelihood of an event.

LOOKING FORWARD

In our future work with inferential statistics, we will use the mean $\bar{x}$ from a random sample to estimate the population parameter μ (Chapter 7) or to make decisions regarding the value of μ (Chapter 8).

◀ Looking Forward

This feature encourages students to pay extra attention to concepts that will be revisited in future chapters.

Real-World Skills

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

UPDATED! Viewpoint Boxes ▶

Viewpoints are activities that use data from actual research studies to help students understand how statistics is applied in the real world. Links to Cengage's Dataset Hub and websites allow students to get hands-on with the data.

VIEWPOINT Loot Boxes in Video Games

Loot boxes are big business in the video game industry, accounting for about $30 billion in revenue annually. Loot boxes, purchased in games for real money, contain a random collection of virtual items for use in the game. Many countries around the world have declared this to be a form of gambling, often targeted at children, and imposed legislation to limit what video game companies can do. At the very least, video games now need to publish the probabilities involved in order to be sold in many countries.

The FIFA series of games from Electronic Arts have a mode that allows players to build a soccer team by buying "card packs" that contain a random selection of player cards, featuring real-world soccer stars, with a range of in-game skill ratings. It is possible to earn some card packs by completing challenges in the game, but to get access to the best cards FIFA players need to spend real money on card packs. The more expensive the card pack, the better the chances of getting a top rated player. However the most expensive card pack has only a 3.2% chance of getting a player in the top tier.

- If we buy 20 packs and count the number of top tier players we get, discuss whether this fits a binomial distribution. What are n, p and q?
- We want to know: If you buy 20 of these packs, what is the probability of getting at least one top tier player? Discuss how you could use SALT to compute this probability.
- We want to know: How many packs must we buy in order to be 80% sure of getting at least one top tier player. Discuss how you might use SALT to find this number. (This type of problem is called a quota problem, and will be discussed in the next section.)

>Tech Notes

Bar graphs, circle graphs, and time-series graphs

TI-84Plus/TI-83Plus/TI-*N*spire (with TI-84Plus keypad) These only graph time series. Place consecutive values 1 through the number of time segments in list L1 and corresponding data in L2. Press **Stat Plot** and highlight an *xy* line plot.

Excel First enter the data into the spreadsheet. Then click the **Insert Tab** and select the type of chart you want to create. A variety of bar graphs, pie charts, and line graphs that can be used as time-series graphs are available. Use the **Design Tab** and + symbol to the right of the graph to access options such as title, axis labels, etc. for your chart. Right clicking the graph or a bar provides other options. The **Format Tab** gives you additional design choices.

Minitab/Minitab Express Use the menu selection **Graph**. Select the desired option and follow the instructions in the dialogue boxes.

SALT After selecting the **Dataset** you wish to work with, go to the **Charts and Graphs** page to select the preferred graph. After selecting the desired option, choose the appropriate **Variable to Graph** from the drop-down menu in the **Settings** bar. If you wish to categorize using a variable included in the data set, choose the appropriate **Group Variable** from the drop-down menu in the **Settings** bar.

◀ REVISED! Tech Notes

Tech Notes give students helpful hints on using TI-84Plus and TI-*N*spire (with TI-84Plus keypad) and TI-83Plus calculators, Microsoft Excel, SALT, Minitab, and Minitab Express to solve a problem.

◀ 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, SALT, Minitab, and Minitab Express.

> USING TECHNOLOGY

Demonstration of the Law of Large Numbers

Computers can be used to simulate experiments. With packages such as Excel, Minitab, and SPSS, programs using random-number generators can be designed (see the *Technology Guide*) to simulate activities such as tossing a die.

The following printouts show the results of the simulations for tossing a die 6, 500, 50,000, 500,000, and 1,000,000 times. Notice how the relative frequencies of the outcomes approach the theoretical probabilities of 1/6 or 0.16667 for each outcome. Consider the following questions based on the simulated data shown below:

- Explain how the result of the simulated data is illustrating the law of large numbers.
- Do you expect the same results every time the simulation is done? Explain.
- Suppose someone gave you a weighted die. Explain how the law or large numbers can help you determine the probability of the die?

Results of Tossing One Die 6 Times

Outcome	Number of Occurrences	Relative Frequency
⚀	0	0.00000
⚁	1	0.16667
⚂	2	0.33333
⚃	0	0.00000
⚄	1	0.16667
⚅	2	0.33333

Results of Tossing One Die 500 Times

Outcome	Number of Occurrences	Relative Frequency
⚀	87	0.17400
⚁	83	0.16600
⚂	91	0.18200
⚃	69	0.13800
⚄	87	0.17400
⚅	83	0.16600

Results of Tossing One Die 50,000 Times

Outcome	Number of Occurrences	Relative Frequency
⚀	8528	0.17056
⚁	8354	0.16708
⚂	8246	0.16492
⚃	8414	0.16828
⚄	8178	0.16356
⚅	8280	0.16560

Results of Tossing One Die 500,000 Times

Outcome	Number of Occurrences	Relative Frequency
⚀	83644	0.16729
⚁	83368	0.16674
⚂	83398	0.16680
⚃	83095	0.16619
⚄	83268	0.16654
⚅	83227	0.16645

Results of Tossing One Die 1,000,000 Times

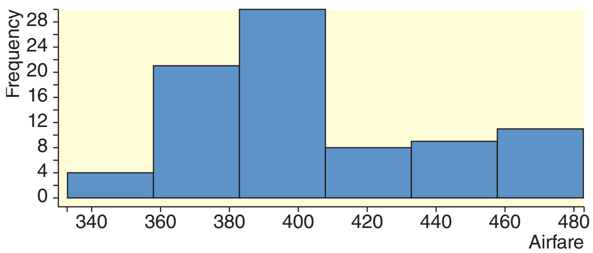
Outcome	Number of Occurrences	Relative Frequency
⚀	166643	0.16664
⚁	166168	0.16617
⚂	167391	0.16739
⚃	165790	0.16579
⚄	167243	0.16724
⚅	166765	0.16677

SALT ▶

SALT (Statistical Analysis and Learning Tool) is integrated into many WebAssign problems to engage students in analyzing, interpreting, and visualizing statistical procedures to understand the meaning behind the data. It is also included in the Tech Notes with instructions, examples, and screenshots.

SALT

The SALT screenshot shows the default histogram created for a dataset using the **Histogram** graph on the **Charts and Graphs** page. Under the **Settings Panel**, you are able to graph the uploaded data by selecting the appropriate variable to graph using the drop down menu under the **Variable to Graph** prompt. Under the **Histogram Settings** you can enter the **Bin/Class Width** and **Starting Point** of your choice and then click the button **Recalculate Bins**. This will generate the necessary histogram based on your specifications.



Data Highlights: Group Projects ▶

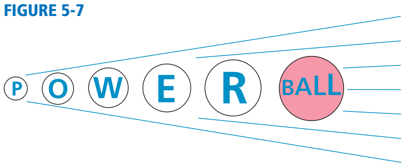
These activities help students synthesize what they've learned by discussing topics, analyzing data, and interpreting results with their classmates.

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. Powerball! Imagine, you could win a jackpot worth at least $40 million. Some jackpots have been worth more than $250 million! Powerball is a multistate lottery. To play Powerball, you purchase a $2 ticket. On the ticket you select five distinct white balls (numbered 1 through 69) and then one red Powerball (numbered 1 through 26). The red Powerball number may be any of the numbers 1 through 26, including any such numbers you selected for the white balls. Every Wednesday and Saturday there is a drawing. If your chosen numbers match those drawn, you win! Figure 5-7 shows all the prizes and the probability of winning each prize and specifies how many numbers on your ticket must match those drawn to win the prize. For updated Powerball data, visit the Powerball web site.

FIGURE 5-7



Match	Approximate Probability	Prize
5 white balls + Powerball	0.0000000034	Jackpot*
5 white balls	0.000000085	$1,000,000
4 white balls + Powerball	0.00000109	$10,000
4 white balls	0.000027	$100
3 white balls + Powerball	0.000069	$100
3 white balls	0.00172	$7
2 white balls + Powerball	0.00143	$7
1 white ball + Powerball	0.01087	$4
0 white balls + Powerball	0.0261	$4
Overall chance of winning	0.0402	

*The Jackpot will be divided equally (if necessary) among multiple winners and is paid in 30 annual installments or in a reduced lump sum.

Making the Jump

Get to the "Aha!" moment faster. *Understandable Statistics: Concepts and Methods* provides the support students need to get there through guidance and example.

PROCEDURE

How to Determine the Number of Outcomes of an Experiment

1. If the experiment consists of a series of stages with various outcomes, use the multiplication rule of counting or a tree diagram.
2. If the outcomes consist of ordered subgroups of r items taken from a group of n items, use the permutations rule, $P_{n,r}$.

$$P_{n,r} = \frac{n!}{(n-r)!} \quad (9)$$

3. If the outcomes consist of nonordered subgroups of r items taken from a group of n items, use the combinations rule, $C_{n,r}$.

$$C_{n,r} = \frac{n!}{r!(n-r)!} \quad (10)$$

◀ Procedures and Requirements

This feature helps reinforce statistical methods by stating the requirements and listing the steps for completing statistical procedures.

UPDATED! Guided Exercises ▶

Guided Exercises walk through statistical methods using real-world contexts and provide worked-out solutions showing students each step in the procedure.

GUIDED EXERCISE 2 **Mean and Trimmed Mean**

Barron's Profiles of American Colleges, 19th edition, lists average class size for introductory lecture courses at each of the profiled institutions. A sample of 20 colleges and universities in California showed class sizes for introductory lecture courses to be

(14)	20	20	20	20	23	25	30	30	30
35	35	35	40	40	42	50	50	80	(80)

(a) Compute a 5% trimmed mean for the sample.

→ The data are already ordered. Since 5% of 20 is 1, we eliminate one data value from the bottom of the list and one from the top. These values are circled in the data set. Then we take the mean of the remaining 18 entries.

$$5\% \text{ trimmed mean} = \frac{\Sigma x}{n} = \frac{625}{18} \approx 34.7$$

(b) Find the median of the original data set.

→ Note that the data are already ordered.

$$\text{Median} = \frac{30+35}{2} = 32.5$$

(c) Find the median of the 5% trimmed data set. Does the median change when you trim the data?

→ The median is still 32.5. Notice that trimming the same number of entries from both ends leaves the middle position of the data set unchanged.

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	Elementary Probability Theory	1.1, 1.2, 2.1, 3.1, 3.2
5	The Binomial Probability	1.1, 1.2, 2.1, 3.1, 3.2, 4.1, 4.2
	Distribution and Related Topics	4.3 useful but not essential
6	Normal Curves and Sampling	
	Distributions (omit 6.6)	1.1, 1.2, 2.1, 3.1, 3.2, 4.1, 4.2, 5.1
	(include 6.6)	also 5.2, 5.3
7	Estimation	
	(omit 7.3 and parts of 7.4)	1.1, 1.2, 2.1, 3.1, 3.2, 4.1, 4.2, 5.1, 6.1, 6.2, 6.3, 6.4, 6.5
	(include 7.3 and all of 7.4)	also 5.2, 5.3, 6.6
8	Hypothesis Testing	
	(omit 8.3 and part of 8.5)	1.1, 1.2, 2.1, 3.1, 3.2, 4.1, 4.2, 5.1, 6.1, 6.2, 6.3, 6.4, 6.5
	(include 8.3 and all of 8.5)	also 5.2, 5.3, 6.6
9	Correlation and Regression	
	(9.1 and 9.2)	1.1, 1.2, 3.1, 3.2
	(9.3 and 9.4)	also 4.1, 4.2, 5.1, 6.1, 6.2, 6.3, 6.4, 6.5, 7.1, 7.2, 8.1, 8.2
10	Chi-Square and *F* Distributions	
	(omit 10.3)	1.1, 1.2, 2.1, 3.1, 3.2, 4.1, 4.2, 5.1, 6.1, 6.2, 6.3, 6.4,
	(include 10.3)	6.5, 7.1 also 8.1
11	Nonparametric Statistics	1.1, 1.2, 2.1, 3.1, 3.2, 4.1, 4.2, 5.1, 6.1, 6.2, 6.3, 6.4, 6.5, 8.1, 8.3

1 Getting Started


Fer Gregory/Shutterstock.com

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)

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.

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?
(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?

LEARNING OBJECTIVES

- **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** 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.

> 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 help you make better decisions. You still have to interpret the statistics and make the decision, but with properly applied statistics you can make an informed decision. Of course, even a properly applied statistical procedure is no more accurate than the data, or facts, on which it is based. 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** 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.

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 left-handed or right-handed.

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*.

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?

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.

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.

LOOKING FORWARD

In later chapters we will use information based on a sample and sample statistics to estimate population parameters (Chapter 7) or make decisions about the value of population parameters (Chapter 8).

For instance, if we have data from *all* the individuals who have climbed Mt. Everest, then we have population data. The proportion of left-handed climbers 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 left-handed climbers in the *sample* is an example of a *statistic*. Note that different samples may have different values for the proportion of left-handed 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.

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 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*.


Kristina Ismulyani/Shutterstock.com

(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

How important is music education in school (K–12)? *The Harris Poll* did an online survey of 2286 adults (aged 18 and older) within the United States. Among the many questions, the survey asked if the respondents agreed or disagreed with the statement, "Learning and habits from music education equip people to be better team players in their careers." In the most recent survey, 71% of the study participants agreed with the statement.

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

→ The individuals are the 2286 adults who participated in the online survey. The variable is the response agree or disagree with the statement that music education equips people to be better team players in their careers.

(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 adults in the United States.

(c) Is the variable qualitative or quantitative?

→ Qualitative—the categories are the two possible responses, agree or disagree with the statement that music education equips people to be better team players in their careers.

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

→ Age or income might be of interest.

(e) Is the proportion of respondents in the sample who agree with the statement regarding music education and effect on careers 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

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.

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.

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

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*

Nick Fox/Shutterstock.com

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.

(b) In a high school graduating class of 319 students, Tatum ranked 25th, Nia ranked 19th, Elias ranked 10th, and Imani ranked 4th, where 1 is the highest rank.

SOLUTION: These data are at the *ordinal* level. Ordering the data clearly makes sense. Elias ranked higher than Nia. Tatum had the lowest rank, and Imani the highest. However, numerical differences in ranks do not have meaning. The difference between Nia's and Tatum's ranks is 6, and this is the same difference that exists between Elias's and Imani's ranks. However, this difference doesn't really mean anything significant. For instance, if you looked at grade point average, Elias and Imani may have had a large gap between their grade point averages, whereas Nia and Tatum 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/tangledlines

(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 Hollis 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, Hollis 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.
(e) The senator surveyed her constituents regarding her 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 her voting record on bills regarding public education.	Ordinal level. Ranks can be ordered, but differences between ranks may vary in meaning.

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.

Keep in mind the fact that statistics from samples will vary from sample to sample, but the parameter for a population is a fixed value that never changes (unless the population changes). Every different sample of a given size from a population will consist of different individual data values, and very likely have different sample statistics. The ways in which sample statistics vary among different samples of the same size will be the focus of our study from Section 6.4 on.

When working with sample data, carefully consider the population from which they are drawn. Observations and analysis of the sample are only applicable to the population from which the sample is drawn.