
Appendix A

Data Files

The data files used in this manual are contained on the Internet. Many are available through the Springer-Verlag web site, in SPSS “.sav” format. The remaining files are available in raw data format on the websites provided; consult Chapter 1 for reading raw data into SPSS. **Note: In the text of this book, we assume that you have retrieved all data, and saved them as SPSS “.sav” files. Thus, all datfiles are referred to as <filename>.sav.**

BEHAVIOR Teacher ratings of 4th grade students' behavior ($n = 2217$).

Variable Name	Variable Description
std9	Student identification number
ssex	Student sex
srace	Student race
1 = white	
2 = black	
3 = Asian	
4 = Hispanic	
5 = Native American	
6 = other	
q1	Pays attention in class
1 = never	
2 = occasionally	
3 = sometimes	
4 = usually	
5 = always	

q2	Completes homework on time
1 = never	
2 = occasionally	
3 = sometimes	
4 = usually	
5 = always	
q3	Works well with others
1 = never	
2 = occasionally	
3 = sometimes	
4 = usually	
5 = always	
q4	Tries to do work well
1 = never	
2 = occasionally	
3 = sometimes	
4 = usually	
5 = always	
q5	Participates in discussions
1 = never	
2 = occasionally	
3 = sometimes	
4 = usually	
5 = always	
q6	Completes seat work
1 = never	
2 = occasionally	
3 = sometimes	
4 = usually	
5 = always	
q7	Thinks school is important
1 = never	
2 = occasionally	
3 = sometimes	
4 = usually	
5 = always	
q8	Does extra work
1 = never	
2 = occasionally	
3 = sometimes	
4 = usually	
5 = always	
q9	Makes effort
1 = never	

2 = occasionally 3 = sometimes 4 = usually 5 = always	
q10	Asks questions
1 = never 2 = occasionally 3 = sometimes 4 = usually 5 = always	
q11	Tries to finish difficult work
1 = never 2 = occasionally 3 = sometimes 4 = usually 5 = always	
q12	Raises hand to talk
1 = never 2 = occasionally 3 = sometimes 4 = usually 5 = always	
q13	Seeks reference material
1 = never 2 = occasionally 3 = sometimes 4 = usually 5 = always	
q14	Discusses outside of class
1 = never 2 = occasionally 3 = sometimes 4 = usually 5 = always	
q15	Attends extracurricular activities
1 = never 2 = occasionally 3 = sometimes 4 = usually 5 = always	

BODYTEMP Body temperature and pulse rate for adults ($n = 130$). Available through *Journal of Statistics Education* Data Archive: <http://www.amstat.org/publications/jse/datasets/normtemp.dat>

Variable Name	Variable Description
temp	Body temperature, degrees Fahrenheit
sex	Sex
0 = female	
1 = male	
pulse	Pulse rate

BOTTLE Daily output of 12 bottle capping machines ($n = 12$). Kruskal, W.H. & Wallis, W.A. (1952). Use of ranks in one-criterion analysis of variance. *Journal of the American Statistical Association*, 47,583–621.

Variable Name	Variable Description
Machine	Machine identification code
Output	Number of bottles capped

CANCER Exposure to radioactive materials and cancer mortality rate ($n = 9$). Fadeley, R.C. (1965). Oregon malignancy pattern physiographically related to Hanford, Washington, radioisotope storage. *Journal of Environmental Health*, 27, 883–897.

Variable Name	Variable Description
expose	Index of exposure
mortalit	Cancer mortality (per 100,000 person years)

CARS Age, color, and owner of cars parked in university parking lot ($n = 64$). Collected in the parking lot of a northeastern university.

Variable Name	Variable Description
color	Color of car
1 = blue	
2 = gray	
3 = red	
4 = black	
5 = green	
6 = white	
7 = brown	
age	Age of car (years)
owner	Car owner
1 = faculty/staff	
2 = student	

CEREAL Nutritional information for breakfast cereals ($n = 77$) Available through the Data and Story Library: <http://lib.stat.cmu.edu/DASL/Datafiles/Cereals.html>

Variable Name	Variable Description
name	Name of cereal
manufac	Manufacturer
1 = American Home Foods	
2 = General Mills	
3 = Kellogg's	
4 = Nabisco	
5 = Post	
6 = Quaker Oats	
7 = Ralston Purina	
type	Type of cereal
1 = cold	
2 = hot	
calories	Calories per serving
protein	Protein grams
fat	Fat grams
sodium	Sodium millimeters
fiber	Fiber
carbo	Carbohydrates
sugar	Sugar
Potass	Potassium
vitamin	Vitamins
shelf	Shelf position in store
1 = bottom	
2 = middle	
3 = top	
weight	Weight (grams)
cups	Cups in serving
rating	Taste rating

CLT 100 random samples of size 50 from uniform distribution ($n = 100$). Data generated by SPSS.

Variable Name	Variable Description
u1	Results of random sampling, time 1
u2	Results of random sampling, time 2
.	
.	
.	
u100	Results of random sampling, time 100

CONFORM Husbands and wives conformity ratings ($n = 20$). Hypothetical data.

Variable Name	Variable Description
husband	Husband's score
wife	Wife's score

CRIME Crime rates per 100,000 for several types of crimes, by state ($n = 50$). Friendly, M. (1999). *Psych6140 Example SAS Programs*. Available: <http://www.psych.yorku.ca/friendly/lab/files/psy6140/examples/factor/pca2.sas>

Variable Name	Variable Description
murder	Murder rate
rape	Rape rate
robbery	Robbery rate
assault	Assault rate
burglary	Burglary rate
larceny	Larceny rate
auto	Automobile crime rate
state	State abbreviation

DEATH Data on number of months before, during, or after birth month that death occurred ($n = 348$). Phillips, D. (1972). Deathday and birthday: An unexpected connection. In J.M. Tanner, *et al.* (Eds.), *Statistics: A guide to the unknown*. San Francisco: Holden Bay.

Variable Name	Variable Description
month	Month of death – month of birth

DELINQ Data on SES, population density, and delinquency for 75 community areas of Chicago ($n = 75$). Hypothetical data suggested by Galle, O.R., Gove, W.R., & McPherson, J.M. (1972). Population density and pathology: What are the relations for man? *Science*, 176, 23–30.

Variable Name	Variable Description
ses	Socioeconomic status (SES)
1 = low	
2 = high	
pop_dens	Population density
1 = low	
2 = high	
delinq	Delinquency
1 = low	
2 = high	

ENROLL Data on school districts, including the racial disproportion in classes for emotionally disturbed children ($n = 26$). U.S. Department of Education, Office for Civil Rights.

Variable Name	Variable Description
enroll	District enrollment
pct_aa	Percentage of students who are African-American
pct_lunch	Percentage of students who pay full-price for lunches
rac_disp	Racial disproportion in classes for emotionally disturbed*

*Positive index indicates that proportion of African-American students is greater than the proportion of white students.

FIRE Data for 28 firefighter applicants ($n = 28$). Buffalo, New York, records.

Variable Name	Variable Description
candnum	Candidate ID number
sex	Sex
1 = male	
2 = female	
race	Race/ethnicity
1 = white	
2 = minority	
stair	Stair climb time (seconds)
body	Body drag time
obstacle	Obstacle course time
agility	Agility score
written	Written score
composite	Composite score

FOOTBALL Data on NFL football games for a recent year ($n = 250$). Publicly kept records.

Variable Name	Variable Description
date	Date of game
aw_tm	Name of away (visiting) team
aw_pt	Number of points away (visiting) team scored
ho_tm	Name of home team
ho_pt	Number of points scored by home team
predptsp	Predicted point spread
predou	Predicted total points
totpnts	Actual total points
actptsp	Actual point spread (winner points – loser points)

favored 1 = home 2 = away winner 1 = home 2 = away actou 0 = even 1 = actual over the predicted 2 = actual under the predicted winby	Favored team Winning team Actual total points compared to predicted total points Points won by
--	--

HOTDOG Nutritional information for different brands of hot dogs ($n = 54$). Available through the Data and Story Library: <http://lib.stat.cmu.edu/DASL/Datafiles/Hotdogs.html>

Variable Name	Variable Description
type 1 = beef 2 = other type of meat 3 = poultry	Type of meat
calories	Calories
sodium	Sodium millimeters

IQ IQ Scores for 23 children ($n = 23$) Anderson, T.W., & Finn, J.D. (1996). *The new statistical analysis of data*. New York: Springer-Verlag.

Variable Name	Variable Description
lang	Language IQ score
nonlang	Nonlanguage IQ score

IQ2 IQ Scores for 24 children ($n = 24$). Anderson, T.W., & Finn, J.D. (1996). *The new statistical analysis of data*. New York: Springer-Verlag.

Variable Name	Variable Description
lang	Language IQ score
nonlang	Nonlanguage IQ score

LIBRARY Size of book collection and number of staff for 22 college libraries ($n = 22$). McGrath, W.E. (1986). Levels of data in the study of library practice: Definition, analysis, inference and explanation. In G. G. Allen & F. C. A. Exon (Eds.), *Research and the practice of librarianship: An international symposium* (pp. 29–40). Perth, Australia: Western Australian Institute of Technology.

Variable Name	Variable Description
volumes	Number of volumes (100,000's)
staff	Number of staff

MOVIES Genre and gross for 100 top movies in 2001 ($n = 100$). Publicly held records.

Variable Name	Variable Description
movie	Movie name
reldate	Release date
opening	Opening week gross (millions)
total	Total gross (millions)
numtheat	Number of theatres in which the movie was released
weekstop	Number of weeks the movie was in the top 60
genre	Genre of movie
1 = thriller/horror	
2 = family	
3 = drama	
4 = comedy	
5 = adventure/fantasy	

NOISE Average highway speed and noise level for 30 sections of highway ($n = 30$). Hypothetical data suggested by Drew, D.R., & Dudek, C.L. (1965). *Investigation of an internal energy model for evaluating freeway level of service*. College Station: Texas A&M University, Texas Transportation Institute.

Variable Name	Variable Description
speed	Acceleration speed (mph)
noise	Noise level

POPULAR Data on elementary school students' goals ($n = 478$). Available through the Data and Story Library: <http://lib.stat.cmu.edu/DASL/Datafiles/PopularKids.html>

Variable Name	Variable Description
gender	Gender
1 = female	
2 = male	
grade	Grade

age	Age
race	Race/ethnicity
1 = white	
2 = other	
urban	School urbanicity
1 = rural	
2 = suburban	
3 = urban	
school	School name
goals	Student goals
1 = make good grades	
2 = be popular	
3 = be good at sports	
grades	Importance of grades for popularity (1 = most; 4 = least)
popular	Importance of sports for popularity (1 = most; 4 = least)
looks	Importance of looks for popularity (1 = most; 4 = least)
money	Importance of money for popularity (1 = most; 4 = least)

READING Reading scores of 30 students before and after second grade ($n = 30$). Records of a second-grade class.

Variable Name	Variable Description
Before	Reading test score before second grade
After	Reading test score after second grade

SLEEP Data on mammals' physical, environmental, and sleep characteristics ($n = 62$). Available: <http://lib.stat.cmu.edu/datasets/sleep>

Variable Name	Variable Description
species	Species' name
bodywt	Body weight (kg)
brainwt	Brain weight (g)
nodream	Non-dreaming sleep (hrs/day)
dream	Dreaming sleep (hrs/day)
totsleep	Total sleep (hrs/day)
lifespan	Life span (years)
gestate	Gestation time (days)
prey	Predation index (1 = min to 5 = max)
sleepexp	Sleep exposure index (1 = least exposed to 5 = most exposed)

danger Danger index – combination of predation and sleep exposure indices (1 = least to 5 = most)

Missing values = -999

SOCMOB Data on family structure and occupation of members ($n = 1156$). Data from: Biblarz, T.J., & Raftery, A.E. Raftery. (1993). The Effects of Family Disruption on Social Mobility. *American Sociological Review*. Data available from StatLib: <http://lib.stat.cmu.edu/datasets/socmob>

Variable Name	Variable Description
Idnum	Family identification number
f_occup	Father's occupation
1 = laborer	
2 = craftsperson	
3 = salesperson	
4 = manager	
5 = professional	
family	Family structure
1 = intact	
2 = non-intact	
race	Race
1 = white	
2 = other	
s_occup	Son's occupation
1 = laborer	
2 = craftsperson	
3 = salesperson	
4 = manager	
5 = professional	

SPIT Data on success of interventions to curb chewing spitting tobacco ($n = 54$). Greene, J.C., Walsh, M.M., & Mosouredis, C. (1994). Report of a pilot study: A program to help major league baseball players quit using spit tobacco. *Journal of the American Dental Association*, 125, 559-567.

Variable	Variable Name
interven	Type of intervention
1 = Minimum	
2 = Extended	
outcome	Outcome of intervention
1 = Subject quit entirely	
2 = Subject tried unsuccessfully to quit	
3 = Subject failed to try to quit	

STEPPING Information on heart rate after stepping exercise, based on differences in stepping frequency and step height ($n = 30$). Available through the Data and Story Library: <http://lib.stat.cmu.edu/DASL/Datafiles/Stepping.html>

Variable Name	Variable Description
Order	Order in study
Block	Subject and experimenter block ID number
Height	Height range of step
0 = low (5.75inches)	
1 = high (11.5 inches)	
Frequency	Frequency of stepping
0 = slow	
1 = medium	
2 = fast/high	
Resthr	Resting heart rate, beats per minute
Hr	Heart rate after exercise, beats per minute

TITANIC Sex, age, and survival outcome for passengers on Titanic. ($n = 2201$). Available through Journal of Statistics Education Data Archive: <http://www.amstat.org/publications/jse/datasets/titanic.dat>

Variable Name	Variable Description
class	Classification - passenger class and crew
0 = crew	
1 = first class	
2 = second class	
3 = third class	
age	Age level
0 = child	
1 = adult	
sex	Sex
0 = female	
1 = male	
survived	Survival status
0 = no	
1 = yes	

WAR Expectations of possibility of war ($n = 597$). Lazarsfeld, P.F., Berelson, B., & Gaudet, H. (1968). *The People's Choice* (3rd edition). New York: Columbia University Press.

Variable Name	Variable Description
June	Response in June 1948
0 = does not expect war	
1 = expects war	

October

Response in October 1948

0 = does not expect war

1 = expects war

WEATHER Average precipitation and temperature on July 2nd for U.S. cities ($n = 78$). Data obtained from the Internet.

Variable Name**Variable Description**

City

Name of city

Temp

Temperature (degrees Fahrenheit)

Precip

Inches of rainfall

WORDS Number of words 18 children memorized based on three different experimental conditions ($n = 18$). Hypothetical data.

Variable**Variable Name**

info_set

Information set

1 = no information

2 = "3 categories"

3 = "6 categories"

words

Number of words memorized

Appendix **B**

Answers to Selected Chapter Exercises

Chapter 1

1.1

- a. 13 variables.
- b. String, 10 characters.

1.2

- a. No, the format is an SPSS data file, not an ASCII data file.
- b. 54 cases.
- c. 2 variables.
- d. No, there are no missing data.

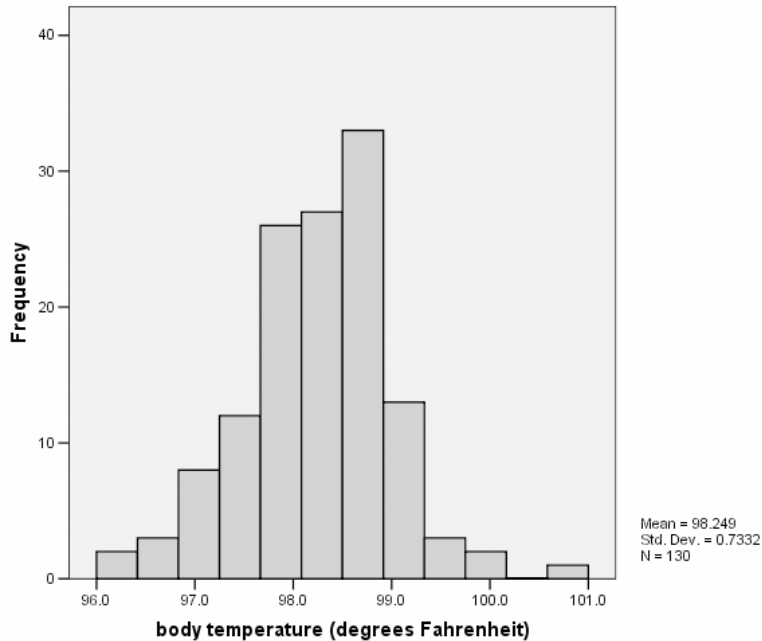
Chapter 2

- a. The frequency distribution of “temp” is given in the table (following page). 10 adults have a body temperature of 98.6°.
- b. 62.3% of the adults in the sample have a body temperature less than 98.6° (that is, 98.5° or less).

body temperature (degrees Fahrenheit)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	96.3	1	.8	.8	.8
	96.4	1	.8	.8	1.5
	96.7	2	1.5	1.5	3.1
	96.8	1	.8	.8	3.8
	96.9	1	.8	.8	4.6
	97.0	1	.8	.8	5.4
	97.1	3	2.3	2.3	7.7
	97.2	3	2.3	2.3	10.0
	97.3	1	.8	.8	10.8
	97.4	5	3.8	3.8	14.6
	97.5	2	1.5	1.5	16.2
	97.6	4	3.1	3.1	19.2
	97.7	3	2.3	2.3	21.5
	97.8	7	5.4	5.4	26.9
	97.9	5	3.8	3.8	30.8
	98.0	11	8.5	8.5	39.2
	98.1	3	2.3	2.3	41.5
	98.2	10	7.7	7.7	49.2
	98.3	5	3.8	3.8	53.1
	98.4	9	6.9	6.9	60.0
	98.5	3	2.3	2.3	62.3
	98.6	10	7.7	7.7	70.0
	98.7	8	6.2	6.2	76.2
	98.8	10	7.7	7.7	83.8
	98.9	2	1.5	1.5	85.4
	99.0	5	3.8	3.8	89.2
	99.1	3	2.3	2.3	91.5
	99.2	3	2.3	2.3	93.8
	99.3	2	1.5	1.5	95.4
	99.4	2	1.5	1.5	96.9
	99.5	1	.8	.8	97.7
	99.9	1	.8	.8	98.5
	100.0	1	.8	.8	99.2
	100.8	1	.8	.8	100.0
	Total	130	100.0	100.0	

- c. The lowest temperature is 96.3°; the highest is 100.8°.
 d. The histogram of “temp” with 12 intervals is listed below:



2.2

- a. There are 14 male firefighter applicants, which is 50% of the sample.
 b. 39.3% of the applicants are minorities.
 c. 64.3% of the women scored below 18 seconds in the stair climb task.
 d. 100% of the men scored below 18 seconds on the stair climb task.
 e. The stem-and-leaf plot of the written test score:

WRITTEN Stem-and-Leaf Plot

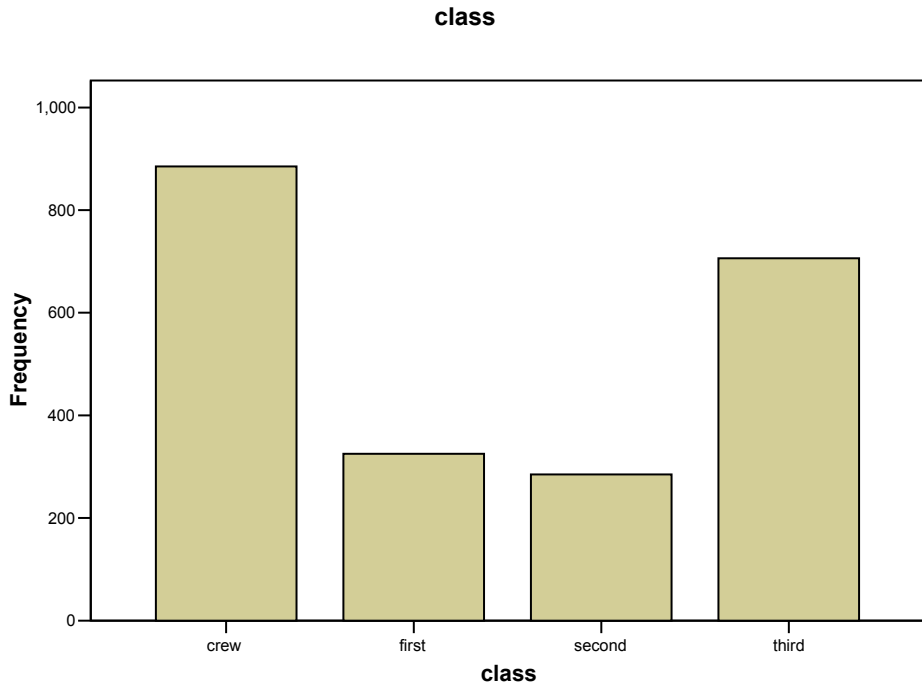
Frequency	Stem & Leaf
7.00	7 . 0011234
9.00	7 . 556678889
3.00	8 . 003
4.00	8 . 6668
3.00	9 . 023
2.00	9 . 58

Stem width: 10.00
 Each leaf: 1 case(s)

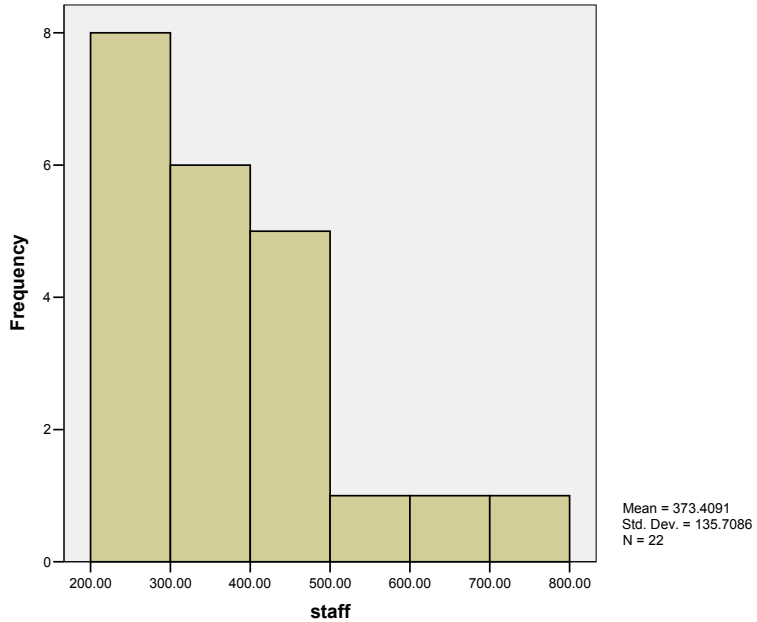
- f. 4 applicants had scored between 85 and 89 on the written test.

2.3

- Bar chart of the “class” variable is listed below. The crew level had the most passengers.
- There were more first-class passengers than second-class passengers.
- 711 passengers survived.

**Chapter 3****3.1**

- mean = 373.4091, median = 346. The differences may be due to the shape of the distribution. Because the mean is larger than the median, the distribution is most likely positively skewed.
- The 10th percentile is 232.2; the 90th is 634.7.
- The histogram is displayed below. The histogram shows the skewness of the distribution. The libraries with over a 600 person staff (specifically, those with 677 and 711 staff members) appear to be outliers.



3.2

- a. The mean is 114.53 seconds.
- b. The mean is 116.53 seconds.
- c. Part (b) is 2 seconds greater than part (a). The general rule is: When you add a constant to each observation in a data set, the mean of the transformed data is equal to the original mean plus the constant.
- d. The mean of the new data is 57.26 seconds, or the original mean divided by 2.

3.3

- a. The mean, median, and mode are contained in the Frequencies table below. Note that only one mode for language IQ is listed. There are five other modes — 94, 95, 99, 102, and 105.

Statistics

		language IQ	nonlanguage IQ
N	Valid	23	23
	Missing	0	0
	Mean	97.57	50.30
	Median	96.00	47.00
	Mode	86 ^a	43

a. Multiple modes exist. The smallest value is shown

- b. The distributions for both of the variables are close to normal, so the mean is the best measure of central tendency for both.

Chapter 4

4.1

- a. The range is 5,200,000 volumes, or 52 (when expressed in 100,000 volumes as is the case in the data file).
- b. The interquartile range (expressed in 100,000) is 19.975. The interquartile range is a better measure of dispersion than is the range in instances when there are outliers in the distribution.
- c. The standard deviation is 14.52879 (100,000 volumes); the variance is 211.086.

4.2

- a. The z-score for *Cats and Dogs* is .16
- b. The results of the Explore procedure are listed below, and indicate that the movie that made the least amount of money in its opening week had a z-score of -1.05 ; the one that made the most had a z-score of 4.00.

Descriptives

			Statistic	Std. Error
Zscore: Opening	Mean		.0000000	.10000000
week gross	95% Confidence Interval for Mean	Lower Bound	-.1984217	
		Upper Bound	.1984217	
	5% Trimmed Mean		-.1130548	
	Median		-.3057197	
	Variance		1.000	
	Std. Deviation		1.000000	
	Minimum		-1.05058	
	Maximum		3.99773	
	Range		5.04831	
	Interquartile Range		.6890511	
	Skewness		1.933	.241
	Kurtosis		3.740	.478

4.3

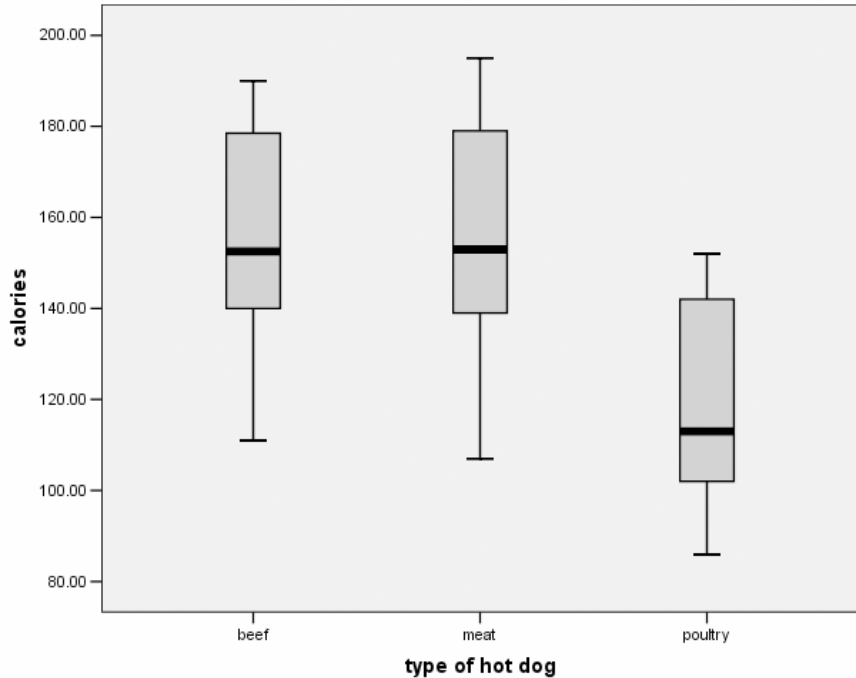
- a. The standard deviation of agility score is 5.72 seconds.
- b. The standard deviation does not change when a constant is subtracted from each time.
- c. The standard deviation is halved when each person's time is halved.

4.4

a. The results of the Explore procedure are displayed below:

Descriptives

type of hot dog			Statistic	Std. Error			
CALORIES	beef	Mean	156.8500	5.06291			
		95% Confidence Interval for Mean	Lower Bound 146.2532 Upper Bound 167.4468				
		5% Trimmed Mean	157.5556				
		Median	152.5000				
		Variance	512.661				
		Std. Deviation	22.64201				
		Minimum	111.00				
		Maximum	190.00				
		Range	79.00				
		Interquartile Range	40.2500				
		Skewness	-.031		.512		
		Kurtosis	-.813		.992		
		meat	Mean		Mean	158.7059	6.12058
					95% Confidence Interval for Mean	Lower Bound 145.7308 Upper Bound 171.6809	
					5% Trimmed Mean	159.5621	
Median	153.0000						
Variance	636.846						
Std. Deviation	25.23580						
Minimum	107.00						
Maximum	195.00						
Range	88.00						
Interquartile Range	42.0000						
Skewness	-.209			.550			
Kurtosis	-.823			1.063			
poultry	Mean			Mean	118.7647	5.46952	
				95% Confidence Interval for Mean	Lower Bound 107.1698 Upper Bound 130.3596		
				5% Trimmed Mean	118.7386		
		Median	113.0000				
		Variance	508.566				
		Std. Deviation	22.55141				
		Minimum	86.00				
		Maximum	152.00				
		Range	66.00				
		Interquartile Range	42.0000				
		Skewness	.025	.550			
		Kurtosis	-1.605	1.063			



- b. The median is 152.5 calories for beef, 153.0 calories for meat, and 113.0 calories for poultry.
- c. The minimum and maximum number of calories for beef hot dogs is 111 calories to 190 calories, respectively.
- d. There are no outliers for poultry hot dogs; the box-and-whisker plot shows that there are no stray points above or below the whiskers.
- e. The meat hot dogs have the most variability (the largest standard deviation).

Chapter 5

5.1

- a. The Pearson correlation is $-.410$

5.2

- a. The Pearson correlation is $.926$
- b. The Spearman correlation is $.833$
- c. Both coefficients indicate a fairly strong, positive association between exposure and mortality.

5.3

The correlation matrix is displayed below:

		Correlations			
		ENROLL	PCT_AA	PCT_LNCH	RAC_DISP
ENROLL	Pearson Correlation	1	-.491*	.089	.204
	Sig. (2-tailed)	.	.011	.666	.319
	N	26	26	26	26
PCT_AA	Pearson Correlation	-.491*	1	-.644**	-.431*
	Sig. (2-tailed)	.011	.	.000	.028
	N	26	26	26	26
PCT_LNCH	Pearson Correlation	.089	-.644**	1	.467*
	Sig. (2-tailed)	.666	.000	.	.016
	N	26	26	26	26
RAC_DISP	Pearson Correlation	.204	-.431*	.467*	1
	Sig. (2-tailed)	.319	.028	.016	.
	N	26	26	26	26

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

- a. The correlation between percentage of students who pay full price for lunches and percent of student who are African-American is largest in magnitude (-.644).
- b. The negative correlation between “enroll” and “pct_aa” indicates that districts with low enrollment tend to have a high percentage of students who are African-American.
- c. Racial disproportion is most highly strongly associated with percentage of students who pay full price for lunch. It is positive and moderately strong (.467).

Chapter 6**6.1**

- a. 62.5% of first class passengers survived; 25.2% of third class passengers survived.
- b. There were 885 crew on board; 24.0% of them survived.
- c. It seems that the first class passengers were more likely to survive than either the crew or third class passengers.

6.2

- a. 21.4% of the applicants were minority females.

- b. White – 9 (or 64.3%) compared to 5 (or 35.7%).
- c. Phi is .073, a very weak association.

6.3

- a. For both boys and girls, making good grades is the most popular goal (selected by 51.8% of females and 51.5% of males). Being good in sports is more important to boys (26.4%) than to girls (12.0%). Similarly, being popular is somewhat more important to girls than to boys (36.3% compared to 22.0%).
- b. The pattern described in part (a) is true in both suburban and urban schools, but not in rural schools. In rural locations, among boys, being good in sports is slightly more popular than is making good grades; and among girls, being popular is almost as important as is making good grades.

Chapter 8

8.1 $z = -2$.

8.2 50%.

8.3 $z = .25$.

8.4 84%.

8.5

- a. 84%.
- b. 95.4%.

Chapter 10

10.1

- a. $t = .482$, $P < .633$. Accept H_0 . Conclude that average speed of vehicles is not different from 35 mph.
- b. No.

10.2

- a. Sample mean = 11.72 pints; 90% confidence interval: (10.80, 12.65)

- b. Yes, reject the null hypothesis because 10 points is not within the confidence interval.
- c. Reject the null hypothesis and conclude that games are won by, on average, more than 10 points.
- d. The two-tailed P value is $P < .002$ so the one-tailed P is $P < .001$.

10.3

- a. $P < .0005$.
- b. Reject at $\alpha = .05$ and at $\alpha = .01$.

10.4

- a. 51.7%
- b. Do not reject the null hypothesis ($P < .493$).

10.5

- a. Yes, wives are more conformist, on average, than their husbands.
- b. The minimum α for rejecting the null hypothesis is .012 (because this is a one-tailed test).

Chapter 11

11.1

- b. Perform a one-tailed test because there is reason to believe that the racial disproportion will be greater in areas that are “low” in terms of percentage of students who pay full price for lunch.
- c. $H_0 : \mu_{\text{low}} \leq \mu_{\text{high}}$ $H_1 : \mu_{\text{low}} > \mu_{\text{high}}$
- d. $t = -1.970$; $P < .060$ for a two-tailed test, so for the one-tailed test it is $P < .030$. Reject H_0 at .05, but not at .01.

11.2

- a. $H_0 : \mu_{\text{home}} \leq \mu_{\text{away}}$ $H_1 : \mu_{\text{home}} > \mu_{\text{away}}$
- b. The P value is .033 (because this is a one tailed test, we divide .066 by 2). Because .033 is not less than .01, we do not reject the null hypothesis, and we conclude that when home teams win, it is not by more points than when away teams win.

11.3

- a. $H_0 : \mu_{\text{students}} = \mu_{\text{faculty}}$ $H_1 : \mu_{\text{students}} \neq \mu_{\text{faculty}}$

- b. The P value is less than .0005, which is less than our significance level, so we concluded that, on average, students drive cars that are older than do faculty (in the sample, the average is 6.25 years for students compared to 3.82 years for faculty).

Chapter 12

12.1

- a. $\chi^2 = 349.915$, $P < .0005$, conclude that class and survival are not independent.
- b. Percentages by class indicate that first class passengers were more likely to survive.

12.2

- a. Sex and views on the importance of money are independent. $\chi^2 = 2.761$, $P < .430$.
- b. Sex and views of the importance of looks are not independent. $\chi^2 = 77.059$, $P < .0005$. The percentages by sex indicate that looks are more important to girls (in the sample, 56.2% of girls rated it most important, compared to 19.4% of boys).

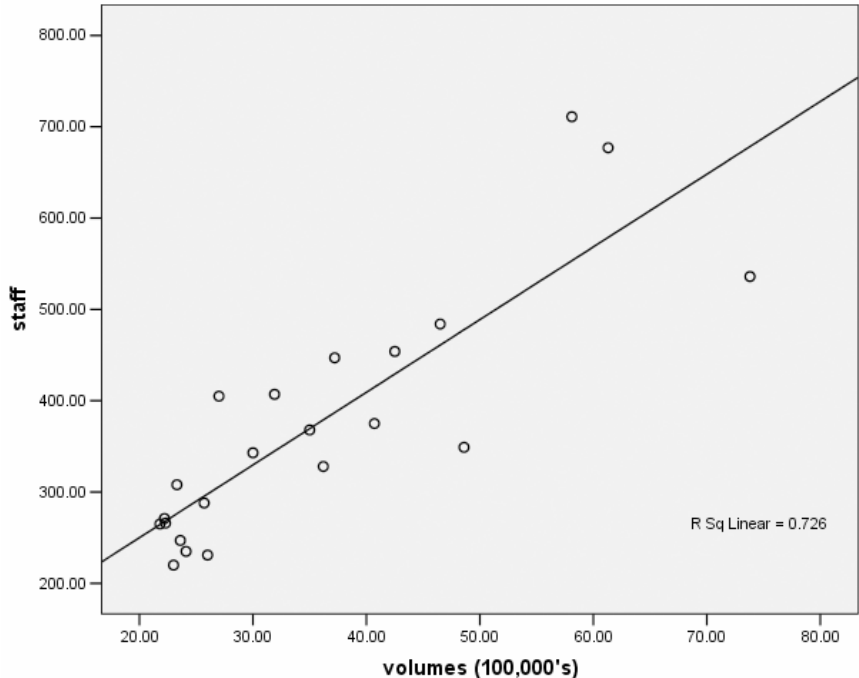
12.3

- a. If we assume that both the variables are ordinal, and that intervention type will predict outcome, then the appropriate measure is Somer's d .
- b. The value of the correlation coefficient is $-.567$, with $P < .0005$. The P value is less than our significance point, so we conclude that the variables are related. The strength of the association is moderate. The negative sign indicates that the extended intervention (which is coded 1 on a 0,1 scale) tends to be related to successful quitting (which is coded 1, one a 1–3 scale).

Chapter 13

13.1

- a. Scatter plot with regression line:



- b. The relationship is positive.
- c. The relationship appears strong.

13.2

- b. The slope, β , is -2.401 , standard error = $.237$, 95% confidence interval for β ($-2.874, -1.928$).
- c. There is a significant relationship between sugar and rating. $t = -10.117$, $P < .0005$. The relationship indicates that cereals with comparatively larger amounts of sugar are rated better tasting (because rating is scored as 1= best tasting).
- d. $r = .760$.
- e. $.577$. 57.7% of the variation in rating of breakfast cereals is attributed to differences in sugar content.

13.3

- c. Yes, there is a significant relationship between gender and agility; men are more agile than women (lower agility scores are superior). This can be discerned in several ways. The simple correlation is $-.766$, $P < .0005$. The

F-test for the Model is $F = 36.943$, $P < .0005$. The t -test for β is $t = -.6078$, $P < .0005$.

- d. Mean for men = $-.7464$; mean for women = 6.0736 . The difference is -6.82 , which is equivalent to the regression coefficient.
- e. The raw coefficient. When the independent variable is a dummy variable, a “one unit increase” is equivalent to moving from one group to the other. Thus, the raw regression coefficient is always equal to the mean difference between groups.

13.4

- a. Yes, the overall $F = 10.259$, and $P < .0005$.
- b. Exposure of the den is related to hours of dream sleep per day controlling for prey index ($t = -2.646$, $P < .011$). The relationship between exposure of the den and amount of dream sleep is such that each unit increase in exposure of an animal’s den (on a 5-point scale) is associated with .397 fewer hours of sleep per day. The likelihood of being preyed upon is not related to amount of dream sleep after controlling for den exposure ($t = -1.016$, $P < .315$).
- c. $R^2 = 30.4\%$.
- d. Overall, there is a relationship among the variables ($F = 7.110$, $P < .001$). The proportion of variance accounted for increases to 31.7% . After controlling for body weight, predation index is not related to dream sleep but den exposure is. The magnitude of the relationship of den exposure increased from the first model, that is, a one-unit increase in exposure is associated with .468 fewer hours of sleep per day.

Chapter 14

14.1

- a. $F = 16.704$, $P < .0005$. Conclude that different types of hot dogs have different average calories.
- b. Poultry hot dogs have fewer calories than either beef ($P < .0005$) or meat ($P < .0005$) hot dogs. There is no significant difference between meat and beef hot dogs ($P < 1.000$). Means in the sample are beef = 156.85 calories, mean = 158.71 calories, poultry = 118.76 calories.
- c. Effect size for poultry with beef is 1.62 standard deviations; for poultry with meat it is 1.70 standard deviations.
- d. $F = 1.778$. Different types of hot dogs do not differ, on average, in amount of sodium.

14.2

- a. $F = 1.867$, $P < .123$. Conclude that genre does not affect the number of weeks movies stay in the Top 60.
- b. Because there is no statistically significant difference, it is not appropriate to conduct post hoc tests.
- c. No post hoc tests were conducted.

Chapter 15**15.1**

- a. 76.81%
- b. Two components had eigenvalues greater than 1.
- c. Because we used the default extraction criteria, the two components with eigenvalues greater than 1 were extracted.
- d. The variables: larceny, auto, burglary and robbery load heavily on the first (rotated) component; murder, assault, and rape load heavily on the second (rotated) component. Thus, the first component may be said to represent “property crime” and the second “personal injury crime.”

15.2

- a. .9486
- b. It decreases to .9417

15.3

The coefficient alpha is .5367, which is somewhat low. If the obstacle course time is deleted, however, the coefficient alpha increases to .9097.

Index

- α , alpha
 - coefficient, 168
 - error, 101
 - intercept, 124
- β , beta, 124
- χ^2 , chi, 117
- γ , gamma, 119
- λ , lambda, 119
- ϕ , phi, 72, 119
- μ , mu, 101
- σ , sigma, 98

- ASCII, 5, 9
- Alternative hypothesis (H_1), 99
- Analysis of variance (ANOVA), 144, 155, 157, 163, 164, 171, 173

- b , least-squares estimate of regression coefficient, 128
- Bar chart, 33
- Bernoulli distribution, 82
- Binomial procedure, 103
- Bonferroni, 147
- Box-and-whisker plot, 56–57

- Central limit theorem, 95, 102

- Change statistics, 137
- Chi-square, 108, 114, 117
- Confidence interval, 110, 115
- Continuity correction, 104
- Correlation coefficient, 69–70, 140
- Correlation matrix, 71–73
- Cross-tabulation, 69
- Cumulative percent, 29

- Data
 - categorical, 27
 - continuous, 27
 - discrete, 27
 - numerical, 27
- Data reduction, 160
- Deletion
 - listwise, 19, 65
 - pairwise, 19, 65
- Descriptive statistics procedure, 28
- Descriptives, 45

- Effect size, 171
- Excel (Microsoft), 9
- Explore procedure, 33

- F statistic, 127, 146
- Frequency distribution, 27

Goodness of Fit, 114

$H (H_0, H_1)$, 99

Help, 5

Histogram, 33, 100

Intercept (y-intercept), 124

Interaction, 151

Kruskal-Wallis, 150

M , median, 42

Main effect, 151

Marginal totals, 71

Maximum, 45

McNemar, 107

Mean

population mean, μ , 101

sample mean, \bar{x} , 44

Minimum, 45

Missing values, 8, 19, 65

Multiple correlation, 128

Nonparametric tests, 103,
107, 114, 150

Null hypothesis (H_0), 99

One-way analysis of
variance, 163–166

Outlier, 52

bivariate, 60

P value, 102, 104, 106,
108, 112, 115, 118, 119,
124, 127, 132, 137, 138,
139, 146, 148, 151

Pearson correlation, 61

Post hoc, 169

Profile plots, 152

r , correlation coefficient,
56

Regression, 142, 147–148,
151–155

Reliability, 187–188

s , sample standard
deviation, 98

Sampling distribution, 91

Scatter plot, 64–68, 140

Skewness, 36

Slope, 124

Spearman-Brown formula,
166

Spearman's rank
correlation, 62

Standard deviation, 55–56

Standard error, 98, 102,
105, 129, 132

Stem-and-leaf, 40

T test, 101

Test statistic, 101

analysis of variance,
146

Chi-square goodness of
fit, 115

independence in two-
way tables, 118

one correlation, 124

one mean, 102

one proportion, 103

regression model, 128

slope of a regression
line, 128

two means, 112

Transform, 5, 12, 63

Type I error, 101

Valid percent, 29, 42

Values labels, 8

Variables

 categorical, 27

 continuous, 27

 dichotomous, 47

 discrete, 27

 numerical, 27

Variance, 50

\bar{x} , sample mean, 44

x bar, 44

y-intercept, 124

z scores, 52