

INTRODUCTION to the **PRACTICE** of **STATISTICS**

EIGHTH EDITION

MOORE ■ McCABE ■ CRAIG



APPLICATIONS

Introduction to the Practice of Statistics presents a wide variety of applications from diverse disciplines. The following list indicates the number of Examples and Exercises related to different fields. Note that some items appear in more than one category.

Examples by Application

Agriculture:

3.19, 4.19, 14.8, 14.9, 15.13, 15.14

Business and consumers:

1.1, 1.2, 1.15, 1.16, 1.23, 1.24, 1.25, 1.26, 1.27, 1.28, 1.30, 1.31, 1.36, 1.47, 1.48, 2.2, 2.8, 2.9, 2.10, 2.11, 2.14, 2.15, 2.16, 2.24, 2.30, 2.31, 2.40, 2.41, 2.43, 3.1, 3.7, 3.9, 3.11, 3.12, 3.13, 3.14, 3.16, 3.17, 3.23, 3.24, 3.25, 3.29, 3.30, 3.31, 3.38, 4.38, 5.2, 5.3, 5.4, 5.5, 5.16, 5.17, 5.18, 5.21, 5.22, 5.23, 5.24, 5.26, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 6.10, 6.11, 6.12, 6.13, 6.14, 6.20, 8.5, 8.6, 10.11, 10.12, 10.13, 11.2, 12.1, 12.2, 12.3, 12.4, 12.5, 12.6, 12.7, 12.8, 12.9, 12.10, 12.11, 12.12, 12.13, 12.14, 12.15, 12.16, 12.17, 12.18, 12.19, 12.20, 12.21, 12.22, 12.23, 12.24, 12.25, 13.1, 13.2, 14.6, 14.10, 14.11, 15.9, 16.1, 16.2, 16.3, 16.10, 17.2, 17.20, 17.21

Demographics and characteristics of people:

1.36, 1.38, 1.39, 5.1, 5.8, 5.12, 5.13, 5.14, 5.19, 5.25, 5.30, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 6.10, 6.11, 6.12, 6.13, 6.14, 6.15, 6.20, 7.1, 7.2, 7.3, 7.13, 12.2, 12.3, 12.4, 12.5, 12.6, 12.7, 12.8, 12.9, 12.10, 12.11, 12.12, 12.13, 12.14, 12.15, 12.16, 12.17, 12.18, 12.19, 12.20, 12.21, 12.22, 12.23, 12.24, 13.7, 13.8

Economics and Finance:

6.31, 7.4, 7.5, 7.6

Education and child development:

1.3, 1.4, 1.5, 1.6, 1.13, 1.17, 1.19, 1.20, 1.29, 1.35, 1.37, 1.40, 1.41, 1.43, 1.44, 1.45, 1.46, 3.2, 3.8, 3.10, 3.20, 3.21, 3.38, 4.18, 4.22, 4.32, 4.33, 4.34, 4.45, 6.3, 6.16, 6.19, 6.27,

7.14, 7.15, 7.18, 11.1, 12.26, 12.27, 15.8, 15.9, 15.10, 16.4, 16.5, 16.6, 16.8, 16.9, 16.11, 16.12, 16.13, 17.19

Environment:

3.33, 6.1, 6.17, 6.18, 6.30

Ethics:

3.34, 3.35, 3.36, 3.37, 3.38, 3.39, 3.40, 3.41, 6.24, 6.25, 6.26

Health and nutrition:

1.11, 1.12, 1.13, 1.21, 1.22, 1.33, 2.1, 2.3, 2.4, 2.5, 2.6, 2.12, 2.13, 2.18, 2.19, 2.20, 2.21, 2.22, 2.23, 2.25, 2.27, 2.28, 2.29, 2.32, 2.34, 2.35, 2.36, 2.37, 2.38, 2.39, 2.43, 2.44, 3.4, 3.5, 3.6, 3.15, 3.18, 3.34, 3.37, 3.39, 4.20, 4.21, 4.26, 4.30, 4.39, 4.41, 5.1, 5.15, 5.19, 5.20, 6.2, 6.15, 6.24, 6.29, 7.16, 7.17, 7.19, 7.20, 7.21, 7.22, 7.23, 8.4, 8.5, 8.6, 8.11, 8.13, 9.8, 9.9, 9.10, 9.11, 9.14, 9.15, 9.16, 9.17, 9.18, 10.1, 10.2, 10.3, 10.4, 10.5, 10.6, 10.7, 10.8, 10.9, 10.10, 10.14, 10.15, 10.16, 10.17, 10.18, 10.19, 10.20, 10.21, 10.22, 10.23, 10.24, 10.25, 10.26, 13.3, 13.4, 13.5, 13.6, 13.7, 13.8, 13.10, 16.14

Humanities and social sciences:

1.32, 3.3, 3.16, 3.27, 3.28, 3.32, 3.40, 3.41, 4.9, 4.10, 4.11, 4.47, 4.48, 5.12, 5.13, 5.14, 6.25, 6.27, 7.7, 7.8, 7.9, 7.12, 8.1, 8.2, 8.3, 8.10, 8.12, 9.1, 9.2, 9.3, 9.4, 9.5, 9.6, 9.7, 9.12, 9.13, 12.3, 13.8, 13.9, 14.1, 14.2, 14.3, 14.4, 14.5, 14.7

International:

1.23, 1.24, 1.25, 1.26, 1.27, 1.36, 1.47, 2.11, 2.15, 2.16, 2.17, 2.26, 3.23, 3.24, 3.25, 5.29, 11.2, 15.6, 16.1, 16.2, 16.3

Law and government data:

3.1, 3.2, 3.3, 3.26, 3.36

Manufacturing, products, and processes:

3.22, 5.11, 5.15, 5.29, 6.17, 6.18, 6.30, 6.32, 10.11, 10.12, 10.13, 17.1, 17.2, 17.3, 17.4, 17.5, 17.6, 17.7, 17.8, 17.9, 17.10, 17.11, 17.12, 17.13, 17.14, 17.15, 17.16, 17.17, 17.18, 17.19, 17.20, 17.21

Science:

1.34, 6.28

Sports and leisure:

1.1, 1.2, 1.40, 1.41, 1.43, 1.44, 1.45, 1.48, 2.17, 3.4, 3.5, 3.6, 4.17, 4.27, 4.31, 4.35, 4.36, 4.37, 4.42, 4.43, 4.46, 6.32, 7.1, 7.2, 7.3, 7.10, 7.11, 8.8, 9.3, 9.4, 9.5, 9.6, 9.7, 9.12, 9.13, 14.6, 14.10, 14.11, 15.1, 15.2, 15.3, 15.4, 15.5, 15.7, 15.11, 15.12, 16.7

Students:

1.3, 1.4, 1.5, 1.6, 1.19, 1.20, 1.35, 2.7, 4.32, 4.33, 4.34, 4.45, 5.1, 5.10, 5.19, 5.28, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 6.10, 6.11, 6.12, 6.13, 6.14, 6.15, 6.16, 6.19, 6.20, 7.1, 7.2, 7.3, 8.7, 8.8, 8.9, 10.1, 10.2, 10.3, 10.4, 10.5, 10.6, 10.7, 10.8, 10.9, 10.10, 10.15, 10.25, 10.26, 11.1, 12.3, 16.4, 16.5, 16.6, 16.8, 16.9, 16.13

Technology and the Internet:

1.1, 1.2, 1.7, 1.8, 1.9, 1.10, 2.26, 3.7, 3.11, 3.12, 3.13, 3.17, 4.26, 4.43, 4.47, 4.48, 5.8, 5.9, 5.27, 5.28, 5.30, 6.28, 7.1, 7.2, 7.3, 7.10, 7.11, 8.1, 8.2, 8.3, 8.10, 8.12, 9.1, 9.2, 12.1, 12.3, 12.4, 12.5, 12.6, 12.7, 12.8, 12.9, 12.10, 12.11, 12.12, 12.13, 12.14, 12.15, 12.16, 12.17, 12.18, 12.19, 12.20, 12.21, 12.22, 12.23, 12.24, 13.1, 13.2, 14.1, 14.2, 14.3, 14.4, 14.5, 14.7, 15.6, 15.7

Exercises by Application

Agriculture:

1.71, 3.64, 3.65, 5.45, 5.46, 5.83, 6.130, 6.131, 12.3, 12.8, 12.13, 13.43, 13.44, 13.45, 13.46, 13.47, 13.48, 13.49, 13.50, 14.31, 14.32, 14.33, 17.75, 17.76

Business and consumers:

1.1, 1.5, 1.8, 1.9, 1.36, 1.37, 1.38, 1.39, 1.47, 1.49, 1.50, 1.53, 1.59, 1.61, 1.62, 1.63, 1.67, 1.68, 1.69, 1.70, 1.77, 1.79, 1.81, 1.156, 1.159, 1.160, 1.161, 1.162, 1.163, 1.174, 2.3, 2.7, 2.10, 2.11, 2.12, 2.13, 2.14, 2.15, 2.16, 2.17, 2.26, 2.27, 2.36, 2.38, 2.39, 2.48, 2.49, 2.51, 2.52, 2.53, 2.54, 2.79, 2.80, 2.91, 2.110, 2.140, 2.149, 2.150, 2.151, 2.152,

2.153, 2.154, 2.155, 2.156, 2.162, 2.163, 2.164, 2.165, 2.166, 2.167, 2.168, 2.172, 3.4, 3.5, 3.9, 3.12, 3.13, 3.15, 3.22, 3.23, 3.27, 3.28, 3.29, 3.30, 3.31, 3.35, 3.39, 3.40, 3.41, 3.42, 3.44, 3.53, 3.59, 3.60, 3.66, 3.76, 3.113, 3.119, 3.123, 3.126, 3.127, 4.38, 4.83, 4.91, 4.92, 4.93, 4.94, 4.121, 4.122, 4.123, 4.124, 4.125, 4.128, 4.141, 4.149, 5.76, 6.1, 6.2, 6.3, 6.6, 6.7, 6.8, 6.22, 6.24, 6.25, 6.29, 6.74, 6.90, 6.91, 6.104, 6.122, 6.124, 6.132, 6.133, 6.140, 7.1, 7.2, 7.3, 7.6, 7.8, 7.9, 7.10, 7.11, 7.16, 7.27, 7.38, 7.42, 7.46, 7.47, 7.71, 7.72, 7.73, 7.77, 7.78, 7.82, 7.83, 7.84, 7.90, 7.91, 7.97, 7.105, 7.106, 7.112, 7.119, 7.120, 7.121, 7.124, 7.130, 7.131, 7.135, 7.140, 7.141, 7.142, 8.1, 8.3, 8.5, 8.6, 8.7, 8.8, 8.9, 8.12, 8.14, 8.15, 8.17, 8.19, 8.30, 8.31, 8.32, 8.37, 8.40, 8.41, 8.44, 8.48, 8.49, 8.50, 8.51, 8.72, 8.84, 8.85, 8.86, 8.101, 9.18, 9.36, 9.37, 9.38, 10.10, 10.11, 10.12, 10.13, 10.14, 10.16, 10.17, 10.18, 10.19, 10.20, 10.21, 10.27, 10.28, 10.30, 11.14, 11.15, 11.16, 11.17, 11.22, 11.23, 11.24, 11.25, 11.26, 11.27, 11.28, 12.14, 12.15, 12.27, 12.35, 12.40, 12.41, 12.42, 12.43, 12.44, 12.69, 13.7, 13.8, 13.13, 13.14, 13.19, 13.20, 13.21, 13.31, 14.3, 14.4, 14.5, 14.6, 14.7, 14.8, 14.11, 14.13, 14.14, 14.26, 14.27, 14.28, 14.34, 14.35, 14.36, 14.41, 14.44, 15.1, 15.2, 15.3, 15.4, 15.5, 15.6, 15.20, 15.21, 15.22, 15.23, 15.24, 15.25, 15.26, 15.27, 15.28, 15.29, 15.32, 15.33, 15.47, 16.1, 16.3, 16.4, 16.9, 16.11, 16.14, 16.15, 16.16, 16.25, 16.30, 16.31, 16.34, 16.47, 16.48, 16.49, 16.53, 16.55, 16.57, 16.61, 16.62, 16.63, 16.64, 16.66, 16.70, 16.73, 16.74, 16.82, 16.83, 16.94, 16.98, 16.99, 17.10, 17.36, 17.38, 17.39, 17.64, 17.69, 17.70, 17.71, 17.72, 17.74, 17.83, 17.84

Demographics and characteristics of people:

1.25, 1.26, 1.27, 1.29, 1.32, 1.33, 2.28, 2.29, 2.30, 2.124, 2.146, 2.147, 2.160, 2.161, 4.56, 5.1, 5.12, 5.13, 5.20, 5.51, 5.52, 5.53, 5.54, 5.55, 5.56, 5.60, 5.62, 5.66, 6.9, 6.10, 6.17, 6.18, 6.23, 6.35, 6.57, 6.66, 7.26, 7.28, 7.144, 8.38, 11.33, 11.34, 11.35, 12.50, 12.63, 13.10, 13.11, 16.11, 16.23

Economics and finance:

5.28, 5.34, 5.58, 5.87, 5.88, 7.87, 10.7, 10.22, 10.25, 10.26, 10.45, 10.46, 10.47, 17.33, 17.34, 17.73

Education and child development:

1.2, 1.3, 1.4, 1.6, 1.7, 1.13, 1.14, 1.17, 1.20, 1.21, 1.22, 1.23, 1.24, 1.43, 1.44, 1.45, 1.48, 1.51, 1.52, 1.54, 1.55, 1.56, 1.57, 1.60, 1.101, 1.102, 1.103, 1.104, 1.105, 1.106, 1.107, 1.108, 1.112, 1.113, 1.114, 1.115, 1.116, 1.117, 1.120, 1.121, 1.132, 1.133, 1.134, 1.135, 1.136, 1.137, 1.138, 1.139, 1.140, 1.141, 1.142, 1.143, 1.173, 1.176, 2.1, 2.21, 2.33, 2.34, 2.45, 2.59, 2.61, 2.73, 2.74, 2.75, 2.76, 2.77, 2.84, 2.87, 2.88, 2.89, 2.90, 2.98, 2.99, 2.100, 2.101, 2.106, 2.113, 2.121, 2.125, 2.137, 2.142, 2.145, 2.157, 2.158, 2.159, 2.175, 2.176, 2.177, 3.10, 3.20, 3.24, 3.25, 3.48, 3.52, 3.82, 3.107, 3.129, 3.132, 3.134, 4.33, 4.46, 4.49, 4.50, 4.99, 4.100, 4.113, 4.114, 4.115, 4.116, 4.117, 4.118, 4.119, 4.120, 4.126, 4.127, 4.142, 4.145, 4.146, 5.22, 5.24, 5.35, 5.64, 5.86, 6.56, 6.64, 6.65, 6.66, 6.71, 6.97, 6.119, 6.123, 6.125, 7.45, 7.50, 7.51, 7.133, 7.137, 7.138, 7.143, 8.59, 8.94, 9.25, 9.26, 9.27, 9.28, 9.33, 9.34, 9.35, 9.39, 9.40, 9.41, 9.42, 9.43, 10.10, 10.11, 10.12, 10.13, 10.14, 10.16, 10.17, 10.18, 10.19, 10.20, 10.21, 10.24, 10.38, 10.44, 10.48, 10.51, 10.52, 10.53, 11.1, 11.3, 11.5, 11.6, 11.13, 11.18, 11.29, 11.30, 11.31, 11.32, 12.13, 12.15, 12.17, 12.20, 12.21, 12.61, 12.62, 12.66, 13.9, 13.32, 13.52, 13.53, 13.54, 13.55, 14.47, 14.48, 14.49, 14.50, 15.8, 15.9, 15.10, 15.11, 15.12, 15.13, 15.16, 15.17, 15.18, 15.19, 15.34, 15.46, 16.18, 16.19, 16.26, 16.27, 16.37, 16.43, 16.45, 16.50, 16.51, 16.52, 16.56, 16.59, 16.65, 16.68, 16.71, 16.72, 16.79, 16.85, 16.86, 16.87, 16.88, 17.79

Environment:

1.30, 1.31, 1.34, 1.35, 1.64, 1.65, 1.66, 1.72, 1.83, 1.88, 1.89, 1.96, 1.100, 1.152, 1.153, 1.156, 2.126, 3.47, 3.49, 3.73, 5.23, 5.45, 5.46, 5.77, 6.35, 6.68, 6.69, 6.116, 6.117, 7.31, 7.77, 7.78, 7.85, 7.86, 7.93, 7.95, 7.96, 7.97, 7.105, 7.106, 7.107, 7.108, 7.110, 7.111, 7.129, 7.132,

8.78, 8.98, 9.55, 10.29, 10.32, 10.33, 10.34, 10.35, 10.36, 10.37, 10.54, 10.55, 10.56, 10.57, 11.20, 11.42, 11.43, 11.44, 11.45, 11.46, 11.47, 11.48, 11.49, 11.50, 11.51, 11.52, 12.14, 12.16, 12.18, 13.12, 13.42, 13.51, 15.24, 15.25, 15.26, 15.27, 15.28, 15.29, 15.35, 15.45, 15.51, 16.28, 16.31, 16.74, 16.75

Ethics:

3.96, 3.97, 3.98, 3.99, 3.100, 3.101, 3.102, 3.103, 3.104, 3.105, 3.106, 3.107, 3.108, 3.109, 3.110, 3.111, 3.112, 3.113, 3.114, 3.115, 3.116, 3.117, 3.119, 3.120, 3.132, 3.133, 3.134, 3.135, 3.136, 3.137, 3.138, 4.113, 4.114, 5.51, 5.52, 5.53, 5.54, 5.55, 5.59, 5.61, 5.81, 6.11, 6.91, 6.92, 6.94, 6.96, 6.100, 6.101, 6.103, 6.104, 7.38, 7.47, 8.33, 8.71, 8.92, 8.93, 8.98, 8.99, 16.98, 16.99

Health and nutrition:

1.15, 1.18, 1.19, 1.73, 1.74, 1.75, 1.82, 1.97, 1.144, 1.145, 1.146, 1.151, 1.157, 1.158, 1.169, 1.170, 1.171, 1.172, 2.2, 2.4, 2.8, 2.18, 2.19, 2.20, 2.35, 2.43, 2.44, 2.62, 2.63, 2.64, 2.66, 2.67, 2.68, 2.69, 2.70, 2.86, 2.92, 2.93, 2.94, 2.95, 2.108, 2.109, 2.115, 2.116, 2.117, 2.118, 2.119, 2.120, 2.123, 2.127, 2.128, 2.129, 2.130, 2.135, 2.136, 2.139, 2.140, 2.141, 2.143, 2.144, 2.170, 2.171, 2.174, 2.178, 3.1, 3.8, 3.11, 3.14, 3.16, 3.18, 3.19, 3.21, 3.26, 3.36, 3.37, 3.38, 3.43, 3.45, 3.75, 3.96, 3.100, 3.101, 3.105, 3.106, 3.109, 3.110, 3.112, 3.114, 3.117, 3.124, 3.130, 3.131, 3.136, 4.29, 4.30, 4.35, 4.39, 4.42, 4.43, 4.44, 4.45, 4.79, 4.82, 4.85, 4.90, 4.107, 4.108, 4.109, 4.111, 4.112, 4.129, 4.130, 4.131, 4.148, 5.21, 5.25, 5.27, 5.32, 5.33, 5.50, 5.56, 5.65, 6.17, 6.18, 6.19, 6.20, 6.23, 6.26, 6.33, 6.34, 6.39, 6.60, 6.67, 6.72, 6.96, 6.99, 6.118, 6.124, 6.125, 6.129, 7.32, 7.34, 7.35, 7.36, 7.37, 7.43, 7.44, 7.48, 7.49, 7.53, 7.61, 7.74, 7.75, 7.76, 7.79, 7.80, 7.89, 7.92, 7.94, 7.98, 7.100, 7.101, 7.103, 7.104, 7.109, 7.121, 7.127, 7.134, 7.136, 7.140, 7.141, 7.142, 8.13, 8.16, 8.18, 8.23, 8.24, 8.42, 8.43, 8.58, 8.59, 8.74, 8.75, 8.76, 8.77, 8.89, 8.90, 9.4, 9.5, 9.6, 9.7, 9.8, 9.14, 9.15, 9.16, 9.17, 9.22, 10.2, 10.5, 10.6, 10.23, 10.42,

10.43, 10.50, 10.58, 10.60, 10.61, 11.21, 11.36, 11.37, 11.38, 11.39, 11.40, 11.41, 12.26, 12.31, 12.32, 12.33, 12.36, 12.39, 12.45, 12.47, 12.48, 12.49, 12.53, 12.68, 13.15, 13.18, 13.22, 13.23, 13.24, 13.25, 13.26, 13.33, 13.36, 13.37, 13.38, 14.15, 14.16, 14.17, 14.18, 14.19, 14.20, 14.21, 14.29, 14.30, 14.37, 14.38, 14.39, 14.40, 14.43, 14.45, 14.46, 15.7, 15.31, 15.36, 15.41, 15.42, 15.43, 15.48, 15.49, 15.50, 15.52, 16.58, 16.67, 16.76, 16.77, 16.78, 16.84, 17.54, 17.78

Humanities and social sciences:

1.76, 1.118, 1.119, 2.107, 2.122, 2.138, 2.139, 3.6, 3.61, 3.63, 3.72, 3.77, 3.78, 3.79, 3.97, 3.98, 3.99, 3.111, 3.115, 3.116, 3.120, 3.133, 3.137, 3.138, 4.12, 4.13, 4.27, 4.28, 4.32, 4.61, 4.65, 4.66, 4.118, 4.119, 4.120, 4.126, 4.127, 4.152, 5.30, 5.31, 5.60, 5.62, 5.66, 6.11, 6.17, 6.18, 6.55, 6.56, 6.62, 6.70, 6.124, 6.140, 7.12, 7.13, 7.14, 7.15, 7.29, 7.33, 7.34, 7.35, 7.55, 7.67, 7.71, 7.90, 7.125, 7.126, 7.128, 7.130, 7.131, 7.139, 8.2, 8.4, 8.12, 8.25, 8.26, 8.28, 8.35, 8.36, 8.71, 8.73, 8.92, 8.93, 8.99, 9.25, 9.26, 9.27, 9.28, 9.40, 9.44, 9.45, 9.49, 10.42, 10.43, 10.50, 10.58, 10.59, 11.19, 11.33, 11.34, 11.35, 12.22, 12.23, 12.24, 12.25, 12.34, 12.36, 12.37, 12.38, 12.46, 12.67, 13.7, 13.8, 13.9, 13.16, 13.17, 13.22, 13.23, 13.24, 13.25, 13.26, 13.27, 13.28, 13.31, 13.35, 13.41, 14.9, 14.10, 15.13, 15.14, 15.15, 15.30, 15.37, 15.38, 15.39, 15.40, 15.45, 16.8, 16.10, 16.40, 16.82, 16.83, 16.89, 16.90, 16.91, 16.92

International:

1.25, 1.26, 1.27, 1.28, 1.29, 1.39, 1.40, 1.41, 1.47, 1.49, 1.61, 1.62, 1.63, 1.73, 1.74, 1.75, 1.119, 1.159, 1.160, 1.161, 1.162, 1.163, 2.15, 2.16, 2.28, 2.29, 2.30, 2.48, 2.53, 2.105, 2.124, 2.146, 2.147, 2.149, 2.150, 2.151, 2.152, 2.153, 2.154, 3.48, 3.66, 4.30, 4.32, 5.52, 5.54, 5.55, 5.59, 5.61, 5.76, 6.72, 6.99, 7.32, 8.23, 8.24, 8.29, 8.60, 8.61, 8.62, 9.41, 9.42, 9.43, 10.39, 10.40, 10.41, 11.33, 11.34, 11.35, 12.37, 12.38, 13.19, 13.20, 13.21, 13.22, 13.23, 13.24, 13.25, 13.26, 14.22, 14.23, 14.35, 14.42, 15.40, 16.3, 16.4,

16.11, 16.16, 16.23, 16.30, 16.34, 16.48, 16.53, 16.57, 16.70

Law and government data:

1.10, 1.30, 1.31, 2.21, 2.160, 2.161, 3.6, 3.67, 3.69, 3.70, 3.111, 3.135, 4.56, 5.29, 5.80, 6.57, 7.113, 9.49, 10.7

Manufacturing, products, and processes:

2.173, 5.17, 5.19, 5.29, 5.50, 5.75, 5.84, 6.30, 6.31, 6.36, 6.39, 6.73, 6.75, 7.27, 7.30, 7.39, 7.40, 7.41, 7.51, 7.52, 10.30, 10.62, 11.53, 11.54, 11.55, 11.56, 11.57, 11.58, 11.59, 11.60, 11.61, 12.51, 12.52, 12.54, 12.55, 12.56, 12.57, 12.58, 12.59, 12.60, 13.7, 13.8, 13.39, 13.40, 15.47, 16.25, 16.37, 16.47, 16.49, 16.84, 17.5, 17.6, 17.7, 17.8, 17.9, 17.11, 17.12, 17.13, 17.14, 17.15, 17.16, 17.17, 17.18, 17.19, 17.20, 17.21, 17.22, 17.29, 17.30, 17.31, 17.32, 17.35, 17.37, 17.40, 17.41, 17.42, 17.43, 17.44, 17.45, 17.46, 17.47, 17.48, 17.49, 17.50, 17.51, 17.57, 17.58, 17.59, 17.60, 17.62, 17.63, 17.65, 17.66, 17.77, 17.80, 17.81, 17.84, 17.86, 17.87, 17.88, 17.89, 17.90, 17.91

Science:

1.100, 2.22, 2.23, 2.46, 2.47, 2.71, 2.72, 2.96, 2.97, 5.38, 5.74, 5.82, 10.7

Sports and leisure:

1.46, 2.19, 2.20, 2.36, 2.37, 2.44, 2.54, 2.67, 2.69, 2.70, 2.91, 2.95, 2.122, 2.148, 3.2, 3.27, 3.28, 3.29, 3.30, 3.31, 3.32, 3.38, 3.50, 3.51, 3.53, 4.5, 4.6, 4.7, 4.9, 4.16, 4.18, 4.34, 4.36, 4.37, 4.58, 4.59, 4.60, 4.80, 4.81, 4.86, 4.89, 4.95, 4.96, 4.97, 4.98, 4.101, 4.136, 4.137, 4.139, 4.143, 4.144, 4.147, 5.11, 5.26, 5.32, 5.33, 5.40, 5.52, 5.54, 5.55, 5.63, 5.79, 5.85, 5.90, 6.9, 6.10, 6.16, 6.27, 6.28, 6.56, 6.72, 6.99, 6.136, 7.7, 7.11, 7.32, 7.134, 8.60, 8.61, 8.62, 8.63, 8.64, 8.65, 8.66, 8.67, 8.82, 8.83, 8.100, 9.44, 9.45, 10.31, 12.13, 12.15, 12.17, 12.19, 14.1, 14.26, 14.34, 15.1, 15.2, 15.3, 15.4, 15.5, 15.6, 15.20, 15.21, 15.22, 15.23, 16.9, 16.12, 16.17, 16.22, 16.24, 16.46, 16.80, 16.81, 16.94, 16.95, 16.96, 16.97, 17.53

Students:

1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.11, 1.12, 1.13, 1.14, 1.17, 1.20, 1.21, 1.22,

1.23, 1.24, 1.28, 1.29, 1.48, 1.51, 1.52, 1.54, 1.55, 1.56, 1.57, 1.60, 1.120, 1.121, 1.164, 1.173, 2.1, 2.5, 2.7, 2.21, 2.32, 2.45, 2.59, 2.73, 2.74, 2.75, 2.76, 2.77, 2.89, 2.90, 2.98, 2.99, 2.100, 2.101, 2.125, 2.137, 2.142, 2.157, 2.158, 2.159, 3.15, 3.20, 3.24, 3.25, 3.52, 3.53, 3.54, 3.55, 3.59, 3.74, 3.79, 3.82, 3.107, 3.129, 4.113, 4.114, 4.115, 4.116, 4.117, 4.122, 4.123, 4.124, 4.125, 4.140, 5.11, 5.14, 5.15, 5.24, 5.36, 5.37, 5.49, 5.51, 5.53, 5.56, 5.64, 5.65, 5.68, 5.87, 6.6, 6.7, 6.8, 6.12, 6.13, 6.14, 6.15, 6.21, 6.22, 6.27, 6.28, 6.38, 6.54, 6.55, 6.56, 6.61, 6.63, 6.71, 6.90, 6.119, 6.121, 6.123, 7.7, 7.26, 7.28, 7.42, 7.68, 7.69, 7.70, 7.113, 7.119, 7.125, 7.126, 7.134, 8.10, 8.11, 8.28, 8.29, 8.30, 8.31, 8.32, 8.33, 8.84, 8.85, 8.86, 8.87, 8.88, 8.94, 9.101, 10.2, 10.5, 10.6, 10.10, 10.11, 10.12, 10.13, 10.14, 10.23, 10.31, 10.42, 10.43, 11.1, 11.3, 11.5, 11.6, 11.13, 11.14, 11.15, 11.16, 11.17, 11.18, 11.21, 12.13, 12.15, 12.17, 12.19, 12.22, 12.43, 12.50, 12.63, 13.52, 13.53, 13.54, 13.55, 15.8, 15.9, 15.10, 15.11, 15.12, 15.13, 15.46, 16.19, 16.26, 16.27, 16.43, 16.45, 16.50, 16.51, 16.52, 16.56, 16.80, 16.81, 16.85, 16.86, 16.87, 16.88, 17.4

Technology and the Internet:

1.1, 1.16, 1.28, 1.29, 1.36, 1.37, 1.38, 1.40, 1.41, 1.165, 1.166, 1.168, 2.5, 2.28, 2.29, 2.30, 2.105, 2.122, 2.137, 2.155, 2.156, 3.5, 3.10, 3.19, 3.22, 3.25, 3.26, 3.27, 3.28, 3.29, 3.30, 3.31, 3.42, 3.44, 3.50, 3.51, 3.64, 3.65, 3.80, 3.107, 3.119, 3.133, 4.11, 4.25, 4.26, 4.27, 4.28, 4.54, 4.55, 5.1, 5.8, 5.11, 5.12, 5.13, 5.16, 5.18, 5.20, 5.37, 5.52, 5.54, 5.55, 5.59, 5.61, 5.64, 5.76, 6.16, 6.57, 6.121, 7.26, 7.28, 7.58, 7.59, 7.68, 7.69, 7.70, 7.81, 7.82, 7.83, 7.128, 8.1, 8.2, 8.3, 8.4, 8.12, 8.15, 8.19, 8.25, 8.33, 8.60, 8.61, 8.62, 8.63, 8.64, 8.65, 8.66, 8.67, 8.72, 8.79, 8.80, 8.81, 8.82, 8.83, 9.1, 9.2, 9.3, 9.10, 9.33, 9.34, 9.35, 12.14, 12.16, 12.18, 12.20, 12.21, 12.22, 12.50, 12.63, 13.19, 13.20, 13.21, 13.34, 14.9, 14.10, 14.11, 14.13, 14.14, 14.22, 14.23, 14.27, 14.28, 14.42, 15.7, 15.30, 15.37, 15.38, 15.39, 16.10, 16.12, 16.17, 16.22, 16.24, 16.46, 16.61, 16.62, 16.63, 16.64, 16.95, 16.96, 16.97

INTRODUCTION
to the **PRACTICE** of
STATISTICS

EIGHTH EDITION

David S. Moore
George P. McCabe
Bruce A. Craig
Purdue University



W. H. Freeman and Company
A Macmillan Higher Education Company

SENIOR PUBLISHER: Ruth Baruth
ACQUISITIONS EDITOR: Karen Carson
MARKETING MANAGER: Steve Thomas
DEVELOPMENTAL EDITOR: Katrina Wilhelm
SENIOR MEDIA EDITOR: Laura Judge
MEDIA EDITOR: Catriona Kaplan
ASSOCIATE EDITOR: Jorge Amaral
ASSISTANT MEDIA EDITOR: Liam Ferguson
PHOTO EDITOR: Cecilia Varas
PHOTO RESEARCHER: Dena Digilio Betz
COVER DESIGNER: Victoria Tomaselli
Text Designer: Patrice Sheridan
PROJECT EDITOR: Elizabeth Geller
ILLUSTRATIONS: Aptara[®], Inc.
PRODUCTION COORDINATOR: Lawrence Guerra
COMPOSITION: Aptara[®], Inc.
PRINTING AND BINDING: QuadGraphics

Library of Congress Control Number: 2013953337

Student Edition Hardcover (packaged with EESEE/CrunchIt! access card):
ISBN-13: 978-1-4641-5893-3
ISBN-10: 1-4641-5893-2

Student Edition Looseleaf (packaged with EESEE/CrunchIt! access card):
ISBN-13: 978-1-4641-5897-1
ISBN-10: 1-4641-5897-5

Instructor Complimentary Copy:
ISBN-13: 978-1-4641-3338-1
ISBN-10: 1-4641-3338-7

© 2014, 2012, 2009, 2006 by W. H. Freeman and Company
All rights reserved

Printed in the United States of America

First printing

W. H. Freeman and Company
41 Madison Avenue
New York, NY 10010
Houndmills, Basingstoke RG21 6XS, England
www.whfreeman.com

BRIEF CONTENTS

To Teachers: About This Book	xiii
To Students: What Is Statistics?	xxiii
About the Authors	xxvii
Data Table Index	xxix
Beyond the Basics Index	xxxi

PART I Looking at Data

CHAPTER 1	Looking at Data— Distributions	1
CHAPTER 2	Looking at Data— Relationships	81
CHAPTER 3	Producing Data	167

PART II Probability and Inference

CHAPTER 4	Probability: The Study of Randomness	231
CHAPTER 5	Sampling Distributions	301
CHAPTER 6	Introduction to Inference	351
CHAPTER 7	Inference for Distributions	417
CHAPTER 8	Inference for Proportions	487

PART III Topics in Inference

CHAPTER 9	Analysis of Two-Way Tables	529
CHAPTER 10	Inference for Regression	563
CHAPTER 11	Multiple Regression	611
CHAPTER 12	One-Way Analysis of Variance	643
CHAPTER 13	Two-Way Analysis of Variance	691
Companion Chapters (on the <i>IPS</i> website www.whfreeman.com/ips8e)		
CHAPTER 14	Logistic Regression	14-1
CHAPTER 15	Nonparametric Tests	15-1
CHAPTER 16	Bootstrap Methods and Permutation Tests	16-1
CHAPTER 17	Statistics for Quality: Control and Capability	17-1
	Tables	T-1
	Answers to Odd-Numbered Exercises	A-1
	Notes and Data Sources	N-1
	Photo Credits	C-1
	Index	I-1

CONTENTS

To Teachers: About This Book	xiii
To Students: What Is Statistics?	xxiii
About the Authors	xxvii
Data Table Index	xxix
Beyond the Basics Index	xxxix

PART I Looking at Data

CHAPTER 1

Looking at Data—Distributions 1

Introduction 1

1.1 Data 2

Key characteristics of a data set 4

Section 1.1 Summary 8

Section 1.1 Exercises 8

1.2 Displaying Distributions with Graphs 9

Categorical variables: bar graphs and pie charts 10

Quantitative variables: stemplots 13

Histograms 15

Data analysis in action: Don't hang up on me 18

Examining distributions 20

Dealing with outliers 21

Time plots 23

Section 1.2 Summary 25

Section 1.2 Exercises 25

1.3 Describing Distributions with Numbers 30

Measuring center: the mean 31

Measuring center: the median 33

Mean versus median 34

Measuring spread: the quartiles 35

The five-number summary and boxplots 37

The $1.5 \times IQR$ rule for suspected outliers 39

Measuring spread: the standard deviation 42

Properties of the standard deviation 44

Choosing measures of center and spread 45

Changing the unit of measurement 45

Section 1.3 Summary 47

Section 1.3 Exercises 48

1.4 Density Curves and Normal Distributions 53

Density curves 55

Measuring center and spread for density curves 56

Normal distributions 58

The 68–95–99.7 rule 59

Standardizing observations 61

Normal distribution calculations 63

Using the standard Normal table 65

Inverse Normal calculations 67

Normal quantile plots 68

Beyond the Basics: Density estimation 71

Section 1.4 Summary 72

Section 1.4 Exercises 72

Chapter 1 Exercises 77

CHAPTER 2

Looking at Data—Relationships 81

Introduction 81

2.1 Relationships 81

Examining relationships 83

Section 2.1 Summary 86

Section 2.1 Exercises 86

2.2 Scatterplots 87

Interpreting scatterplots 89

The log transformation 93

Adding categorical variables to scatterplots 94

Beyond the Basics: Scatterplot smoothers 94

Categorical explanatory variables 97

Section 2.2 Summary 97

Section 2.2 Exercises 98

2.3 Correlation 103

The correlation r 103

Properties of correlation 105

Section 2.3 Summary 107

Section 2.3 Exercises 107

2.4 Least-Squares Regression 109

Fitting a Line to Data 110

Prediction 112

Least-squares regression 113

Interpreting the regression line 116

Facts about least-squares regression	118	How to randomize	181
Correlation and regression	119	Cautions about experimentation	185
Another view of r^2	120	Matched pairs designs	186
Section 2.4 Summary	121	Block designs	187
Section 2.4 Exercises	122	Section 3.2 Summary	188
2.5 Cautions about Correlation and Regression	126	Section 3.2 Exercises	189
Residuals	126	3.3 Sampling Design	192
Outliers and influential observations	130	Simple random samples	194
Beware of the lurking variable	132	Stratified random samples	196
Beware of correlations based on averaged data	134	Multistage random samples	197
Beware of restricted ranges	134	Cautions about sample surveys	198
Beyond the Basics: Data mining	135	Section 3.3 Summary	201
Section 2.5 Summary	136	Section 3.3 Exercises	201
Section 2.5 Exercises	136	3.4 Toward Statistical Inference	205
2.6 Data Analysis for Two-Way Tables	139	Sampling variability	207
The two-way table	139	Sampling distributions	208
Joint distribution	141	Bias and variability	210
Marginal distributions	142	Sampling from large populations	212
Describing relations in two-way tables	143	Why randomize?	213
Conditional distributions	144	Beyond the Basics: Capture-recapture sampling	214
Simpson's paradox	146	Section 3.4 Summary	215
Section 2.6 Summary	148	Section 3.4 Exercises	215
Section 2.6 Exercises	148	3.5 Ethics	217
2.7 The Question of Causation	152	Institutional review boards	219
Explaining association	152	Informed consent	220
Establishing causation	154	Confidentiality	220
Section 2.7 Summary	156	Clinical trials	222
Section 2.7 Exercises	156	Behavioral and social science experiments	224
Chapter 2 Exercises	157	Section 3.5 Summary	226
		Section 3.5 Exercises	226
		Chapter 3 Exercises	228

CHAPTER 3

Producing Data 167

Introduction 167

3.1 Sources of Data 168

Anecdotal data	168
Available data	169
Sample surveys and experiments	171

Section 3.1 Summary 174

Section 3.1 Exercises 174

3.2 Design of Experiments 175

Comparative experiments	178
Randomization	179
Randomized comparative experiments	181

PART II Probability and Inference

CHAPTER 4

Probability: The Study of Randomness 231

Introduction 231

4.1 Randomness 231

The language of probability	233
Thinking about randomness	234
The uses of probability	235
Section 4.1 Summary	235
Section 4.1 Exercises	236

4.2 Probability Models	236	The central limit theorem	307
Sample spaces	237	A few more facts	313
Probability rules	240	Beyond the Basics: Weibull distributions	315
Assigning probabilities: finite number of outcomes	242	Section 5.1 Summary	316
Assigning probabilities: equally likely outcomes	243	Section 5.1 Exercises	316
Independence and the multiplication rule	244	5.2 Sampling Distributions for Counts and Proportions	320
Applying the probability rules	248	The binomial distributions for sample counts	322
Section 4.2 Summary	248	Binomial distributions in statistical sampling	324
Section 4.2 Exercises	249	Finding binomial probabilities	325
4.3 Random Variables	252	Binomial mean and standard deviation	328
Discrete random variables	253	Sample proportions	329
Continuous random variables	256	Normal approximation for counts and proportions	331
Normal distributions as probability distributions	259	The continuity correction	335
Section 4.3 Summary	260	Binomial formula	336
Section 4.3 Exercises	261	The Poisson distributions	339
4.4 Means and Variances of Random Variables	263	Section 5.2 Summary	343
The mean of a random variable	264	Section 5.2 Exercises	344
Statistical estimation and the law of large numbers	267	Chapter 5 Exercises	349
Thinking about the law of large numbers	269	CHAPTER 6	
Beyond the Basics: More laws of large numbers	270	Introduction to Inference	351
Rules for means	271	Introduction	351
The variance of a random variable	273	Overview of inference	352
Rules for variances and standard deviations	275	6.1 Estimating with Confidence	353
Section 4.4 Summary	279	Statistical confidence	354
Section 4.4 Exercises	280	Confidence intervals	356
4.5 General Probability Rules	282	Confidence interval for a population mean	358
General addition rules	283	How confidence intervals behave	362
Conditional probability	286	Choosing the sample size	363
General multiplication rules	289	Some cautions	365
Tree diagrams	290	Beyond the Basics: The bootstrap	367
Bayes's rule	292	Section 6.1 Summary	368
Independence again	293	Section 6.1 Exercises	368
Section 4.5 Summary	294	6.2 Tests of Significance	372
Section 4.5 Exercises	294	The reasoning of significance tests	372
Chapter 4 Exercises	297	Stating hypotheses	374
		Test statistics	375
CHAPTER 5		P-values	377
Sampling Distributions	301	Statistical significance	378
Introduction	301	Tests for a population mean	382
5.1 The Sampling Distribution of a Sample Mean	303	Two-sided significance tests and confidence intervals	386
The mean and standard deviation of \bar{x}	305	The P-value versus a statement of significance	387
		Section 6.2 Summary	390
		Section 6.2 Exercises	390

6.3 Use and Abuse of Tests	394	7.3 Other Topics in Comparing Distributions	473
Choosing a level of significance	395	Inference for population spread	474
What statistical significance does not mean	396	The F test for equality of spread	474
Don't ignore lack of significance	397	Robustness of Normal inference procedures	477
Statistical inference is not valid for all sets of data	398	The power of the two-sample t test	477
Beware of searching for significance	399	Section 7.3 Summary	479
Section 6.3 Summary	400	Section 7.3 Exercises	480
Section 6.3 Exercises	400	Chapter 7 Exercises	481
6.4 Power and Inference as a Decision	402	CHAPTER 8	
Power	402	Inference for Proportions	487
Increasing the power	405	Introduction	487
Inference as decision	406	8.1 Inference for a Single Proportion	488
Two types of error	407	Large-sample confidence interval for a single proportion	489
Error probabilities	408	Beyond the Basics: The plus four confidence interval for a single proportion	493
The common practice of testing hypotheses	410	Significance test for a single proportion	495
Section 6.4 Summary	411	Choosing a sample size	500
Section 6.4 Exercises	411	Section 8.1 Summary	503
Chapter 6 Exercises	413	Section 8.1 Exercises	504
CHAPTER 7		8.2 Comparing Two Proportions	508
Inference for Distributions	417	Large-sample confidence interval for a difference in proportions	509
Introduction	417	Beyond the Basics: Plus four confidence interval for a difference in proportions	514
7.1 Inference for the Mean of a Population	418	Significance test for a difference in proportions	516
The t distributions	418	Beyond the Basics: Relative risk	520
The one-sample t confidence interval	420	Section 8.2 Summary	521
The one-sample t test	422	Section 8.2 Exercises	522
Matched pairs t procedures	429	Chapter 8 Exercises	525
Robustness of the t procedures	432		
The power of the t test	434		
Inference for non-Normal populations	436		
Section 7.1 Summary	440		
Section 7.1 Exercises	441		
7.2 Comparing Two Means	447	PART III Topics in Inference	
The two-sample z statistic	448	CHAPTER 9	
The two-sample t procedures	450	Analysis of Two-Way Tables	529
The two-sample t confidence interval	451	Introduction	529
The two-sample t significance test	454	9.1 Inference for Two-Way Tables	530
Robustness of the two-sample procedures	455	The hypothesis: no association	536
Inference for small samples	457	Expected cell counts	537
Software approximation for the degrees of freedom	460	The chi-square test	537
The pooled two-sample t procedures	461	Computations	540
Section 7.2 Summary	466		
Section 7.2 Exercises	467		

Computing conditional distributions	541	11.2 A Case Study	619
The chi-square test and the z test	544	Preliminary analysis	619
Models for two-way tables	545	Relationships between pairs of variables	621
Beyond the Basics: Meta-analysis	548	Regression on high school grades	623
Section 9.1 Summary	550	Interpretation of results	624
9.2 Goodness of Fit	551	Residuals	625
Section 9.2 Summary	556	Refining the model	625
Chapter 9 Exercises	557	Regression on SAT scores	627
		Regression using all variables	628
		Test for a collection of regression coefficients	630
CHAPTER 10		Beyond the Basics: Multiple logistic regression	631
Inference for Regression	563	Chapter 11 Summary	633
Introduction	563	Chapter 11 Exercises	634
10.1 Simple Linear Regression	564	CHAPTER 12	
Statistical model for linear regression	564	One-Way Analysis of Variance	643
Data for simple linear regression	566	Introduction	643
Estimating the regression parameters	568	12.1 Inference for One-Way Analysis of Variance	644
Confidence intervals and significance tests	574	Data for one-way ANOVA	644
Confidence intervals for mean response	576	Comparing means	645
Prediction intervals	578	The two-sample t statistic	647
Transforming variables	580	An overview of ANOVA	647
Beyond the Basics: Nonlinear regression	582	The ANOVA model	651
Section 10.1 Summary	584	Estimates of population parameters	653
10.2 More Detail about Simple Linear Regression	585	Testing hypotheses in one-way ANOVA	655
Analysis of variance for regression	586	The ANOVA table	657
The ANOVA F test	588	The F test	660
Calculations for regression inference	590	12.2 Comparing the Means	663
Inference for correlation	597	Contrasts	663
Section 10.2 Summary	599	Multiple comparisons	668
Chapter 10 Exercises	600	Software	673
		Power	675
		Chapter 12 Summary	677
		Chapter 12 Exercises	678
CHAPTER 11		CHAPTER 13	
Multiple Regression	611	Two-Way Analysis of Variance	691
Introduction	611	Introduction	691
11.1 Inference for Multiple Regression	612	13.1 The Two-Way ANOVA Model	692
Population multiple regression equation	612	Advantages of two-way ANOVA	692
Data for multiple regression	613	The two-way ANOVA model	696
Multiple linear regression model	614	Main effects and interactions	697
Estimation of the multiple regression parameters	615		
Confidence intervals and significance tests for regression coefficients	616		
ANOVA table for multiple regression	617		
Squared multiple correlation R^2	618		

13.2 Inference for Two-Way ANOVA	702	15.3 The Kruskal-Wallis Test	15-28
The ANOVA table for two-way ANOVA	702	Hypotheses and assumptions	15-29
Chapter 13 Summary	708	The Kruskal-Wallis test	15-29
Chapter 13 Exercises	708	Section 15.3 Summary	15-33
		Section 15.3 Exercises	15-33
		Chapter 15 Exercises	15-35
		Chapter 15 Notes and Data Sources	15-36
Companion Chapters			
(on the <i>IPS</i> website www.whfreeman.com/ips8e)			
CHAPTER 14			
Logistic Regression	14-1	CHAPTER 16	
Introduction	14-1	Bootstrap Methods and Permutation Tests	16-1
14.1 The Logistic Regression Model	14-2	Introduction	16-1
Binomial distributions and odds	14-2	Software	16-2
Odds for two groups	14-3	16.1 The Bootstrap Idea	16-3
Model for logistic regression	14-5	The big idea: resampling and the bootstrap distribution	16-4
Fitting and interpreting the logistic regression model	14-6	Thinking about the bootstrap idea	16-8
		Using software	16-9
14.2 Inference for Logistic Regression	14-9	Section 16.1 Summary	16-10
Confidence intervals and significance tests	14-10	Section 16.1 Exercises	16-11
Multiple logistic regression	14-16	16.2 First Steps in Using the Bootstrap	16-13
Chapter 14 Summary	14-19	Bootstrap t confidence intervals	16-14
Chapter 14 Exercises	14-20	Bootstrapping to compare two groups	16-17
Chapter 14 Notes and Data Sources	14-27	Beyond the Basics: The bootstrap for a scatterplot smoother	16-20
		Section 16.2 Summary	16-22
		Section 16.2 Exercises	16-22
CHAPTER 15			
Nonparametric Tests	15-1	16.3 How Accurate Is a Bootstrap Distribution?	16-24
Introduction	15-1	Bootstrapping small samples	16-27
15.1 The Wilcoxon Rank Sum Test	15-3	Bootstrapping a sample median	16-29
The rank transformation	15-4	Section 16.3 Summary	16-30
The Wilcoxon rank sum test	15-5	Section 16.3 Exercises	16-31
The Normal approximation	15-7	16.4 Bootstrap Confidence Intervals	16-32
What hypotheses does Wilcoxon test?	15-9	Bootstrap percentile confidence intervals	16-32
Ties	15-10	A more accurate bootstrap confidence interval: BCa	16-34
Rank, t , and permutation tests	15-13	Confidence intervals for the correlation	16-36
Section 15.1 Summary	15-15	Section 16.4 Summary	16-38
Section 15.1 Exercises	15-15	Section 16.4 Exercises	16-38
15.2 The Wilcoxon Signed Rank Test	15-18	16.5 Significance Testing Using Permutation Tests	16-42
The Normal approximation	15-22	Using software	16-46
Ties	15-23		
Testing a hypothesis about the median of a distribution	15-25		
Section 15.2 Summary	15-25		
Section 15.2 Exercises	15-25		

Permutation tests in practice	16-46	Setting up control charts	17-26
Permutation tests in other settings	16-49	Comments on statistical control	17-31
Section 16.5 Summary	16-52	Don't confuse control with capability!	17-34
Section 16.5 Exercises	16-53	Section 17.2 Summary	17-35
Chapter 16 Exercises	16-56	Section 17.2 Exercises	17-36
Chapter 16 Notes and Data Sources	16-58	17.3 Process Capability Indexes	17-41
CHAPTER 17		The capability indexes C_p and C_{pk}	17-43
Statistics for Quality: Control and Capability	17-1	Cautions about capability indexes	17-46
Introduction	17-1	Section 17.3 Summary	17-48
Use of data to assess quality	17-2	Section 17.3 Exercises	17-48
17.1 Processes and Statistical Process Control	17-3	17.4 Control Charts for Sample Proportions	17-52
Describing processes	17-3	Control limits for p charts	17-53
Statistical process control	17-6	Section 17.4 Summary	17-57
\bar{x} charts for process monitoring	17-8	Section 17.4 Exercises	17-57
s charts for process monitoring	17-12	Chapter 17 Exercises	17-58
Section 17.1 Summary	17-17	Chapter 17 Notes and Data Sources	17-60
Section 17.1 Exercises	17-18	Tables	T-1
17.2 Using Control Charts	17-22	Answers to Odd-Numbered Exercises	A-1
\bar{x} and R charts	17-23	Notes and Data Sources	N-1
Additional out-of-control rules	17-24	Photo Credits	C-1
		Index	I-1

Statistics is the science of data. *Introduction to the Practice of Statistics (IPS)* is an introductory text based on this principle. We present methods of basic statistics in a way that emphasizes working with data and mastering statistical reasoning. *IPS* is elementary in mathematical level but conceptually rich in statistical ideas. After completing a course based on our text, we would like students to be able to think objectively about conclusions drawn from data and use statistical methods in their own work.

In *IPS* we combine attention to basic statistical concepts with a comprehensive presentation of the elementary statistical methods that students will find useful in their work. *IPS* has been successful for several reasons:

1. *IPS* examines the nature of modern statistical practice at a level suitable for beginners. We focus on the production and analysis of data as well as the traditional topics of probability and inference.
2. *IPS* has a logical overall progression, so data production and data analysis are a major focus, while inference is treated as a tool that helps us draw conclusions from data in an appropriate way.
3. *IPS* presents data analysis as more than a collection of techniques for exploring data. We emphasize systematic ways of thinking about data. Simple principles guide the analysis: always plot your data; look for overall patterns and deviations from them; when looking at the overall pattern of a distribution for one variable, consider shape, center, and spread; for relations between two variables, consider form, direction, and strength; always ask whether a relationship between variables is influenced by other variables lurking in the background. We warn students about pitfalls in clear cautionary discussions.
4. *IPS* uses real examples to drive the exposition. Students learn the technique of least-squares regression and how to interpret the regression slope. But they also learn the conceptual ties between regression and correlation and the importance of looking for influential observations.
5. *IPS* is aware of current developments both in statistical science and in teaching statistics. Brief optional Beyond the Basics sections give quick overviews of topics such as density estimation, scatterplot smoothers, data mining, nonlinear regression, and meta-analysis. Chapter 16 gives an elementary introduction to the bootstrap and other computer-intensive statistical methods.

The title of the book expresses our intent to introduce readers to statistics as it is used in practice. Statistics in practice is concerned with drawing conclusions from data. We focus on problem solving rather than on methods that may be useful in specific settings.

GAISE The College Report of the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Project (<http://www.amstat.org/education/gaise/>) was funded by the American Statistical Association to make recommendations for how introductory statistics courses should be taught. This report contains many interesting teaching suggestions and we strongly recommend

that you read it. The philosophy and approach of *IPS* closely reflect the GAISE recommendations. Let's examine each of the recommendations in the context of *IPS*.

1. **Emphasize statistical literacy and develop statistical thinking.** Through our experiences as applied statisticians, we are very familiar with the components that are needed for the appropriate use of statistical methods. We focus on collecting and finding data, evaluating the quality of data, performing statistical analyses, and drawing conclusions. In examples and exercises throughout the text, we emphasize putting the analysis in the proper context and translating numerical and graphical summaries into conclusions.
2. **Use real data.** Many of the examples and exercises in *IPS* include data that we have obtained from collaborators or consulting clients. Other data sets have come from research related to these activities. We have also used the Internet as a data source, particularly for data related to social media and other topics of interest to undergraduates. With our emphasis on real data, rather than artificial data chosen to illustrate a calculation, we frequently encounter interesting issues that we explore. These include outliers and nonlinear relationships. All data sets are available from the text website.
3. **Stress conceptual understanding rather than mere knowledge of procedures.** With the software available today, it is very easy for almost anyone to apply a wide variety of statistical procedures, both simple and complex, to a set of data. Without a firm grasp of the concepts, such applications are frequently meaningless. By using the methods that we present on real sets of data, we believe that students will gain an excellent understanding of these concepts. Our emphasis is on the input (questions of interest, collecting or finding data, examining data) and the output (conclusions) for a statistical analysis. Formulas are given only where they will provide some insight into concepts.
4. **Foster active learning in the classroom.** As we mentioned above, we believe that statistics is exciting as something to do rather than something to talk about. Throughout the text we provide exercises in Use Your Knowledge sections that ask the students to perform some relatively simple tasks that reinforce the material just presented. Other exercises are particularly suited to being worked and discussed within a classroom setting.
5. **Use technology for developing concepts and analyzing data.** Technology has altered statistical practice in a fundamental way. In the past, some of the calculations that we performed were particularly difficult and tedious. In other words, they were not fun. Today, freed from the burden of computation by software, we can concentrate our efforts on the big picture: what questions are we trying to address with a study and what can we conclude from our analysis?
6. **Use assessments to improve and evaluate student learning.** Our goal for students who complete a course based on *IPS* is that they are able to design and carry out a statistical study for a project in their capstone course or other setting. Our exercises are oriented toward this goal. Many ask about the design of a statistical study and the collection of data. Others ask for a paragraph summarizing the results of an analysis. This recommendation includes the use of projects, oral presentations, article

critiques, and written reports. We believe that students using this text will be well prepared to undertake these kinds of activities. Furthermore, we view these activities not only as assessments but also as valuable tools for learning statistics.

Teaching Recommendations We have used *IPS* in courses taught to a variety of student audiences. For general undergraduates from mixed disciplines, we recommend covering Chapters 1 to 8 and Chapters 9, 10, or 12. For a quantitatively strong audience—sophomores planning to major in actuarial science or statistics—we recommend moving more quickly. Add Chapters 10 and 11 to the core material in Chapters 1 to 8. In general, we recommend de-emphasizing the material on probability because these students will take a probability course later in their program. For beginning graduate students in such fields as education, family studies, and retailing, we recommend that the students read the entire text (Chapters 11 and 13 lightly), again with reduced emphasis on Chapter 4 and some parts of Chapter 5. In all cases, beginning with data analysis and data production (Part I) helps students overcome their fear of statistics and builds a sound base for studying inference. We believe that *IPS* can easily be adapted to a wide variety of audiences.

The Eighth Edition: What's New?

- **Text Organization** Each section now begins with the phrase “When you complete this section, you will be able to” followed by a bulleted list of behavioral objectives that the students should be able to master. Exercises that focus on these objectives appear at the beginning of the section exercises. The long introduction to **Chapter 1** has been replaced by a short introduction and a new section titled “Data,” which gives an overview of the basic ideas on the key characteristics of a set of data. The same approach has been taken with **Chapters 2 and 3**, which now have new sections titled “Relationships” and “Sources of Data,” respectively. A short introduction to the Poisson distribution has been added to **Section 5.2**. **Sections 9.1 and 9.2** have been combined with a more concise presentation of the material on computation and models from Section 5.2 of the seventh edition. In **Chapter 16**, the use of S-PLUS software has been replaced by R. Sections previously marked as **optional** are no longer given this designation. We have found that instructors make a variety of choices regarding what to include in their courses. General guidelines for different types of students are given in the Teaching Recommendations paragraph above.
- **Design** A new design incorporates colorful, revised figures throughout to aid the students’ understanding of text material. Photographs related to chapter examples and exercises make connections to real-life applications and provide a visual context for topics. More figures with software output have been included.
- **Exercises and Examples** Over 50% of the exercises are new or revised. There are more than 1700 exercises, a slight increase over the total in the seventh edition. To maintain the attractiveness of the examples to students, we have replaced or updated a large number of them. Over 35% of the 422 examples are new or revised. A list of exercises and examples categorized by application area is provided on the inside of the front cover.

In addition to the new eighth edition enhancements, *IPS* has retained the successful pedagogical features from previous editions:



- **Look Back** At key points in the text, Look Back margin notes direct the reader to the first explanation of a topic, providing page numbers for easy reference.



- **Caution** Warnings in the text, signaled by a caution icon, help students avoid common errors and misconceptions.



- **Challenge Exercises** More challenging exercises are signaled with an icon. Challenge exercises are varied: some are mathematical, some require open-ended investigation, and others require deeper thought about the basic concepts.



- **Applets** Applet icons are used throughout the text to signal where related interactive statistical applets can be found on the *IPS* website.



- **Use Your Knowledge Exercises** We have found these to be a very useful learning tool. Therefore, we have increased the number and variety of these exercises. These exercises are listed, with page numbers, before the section-ending exercises.

Acknowledgments

We are pleased that the first seven editions of *Introduction to the Practice of Statistics* have helped to move the teaching of introductory statistics in a direction supported by most statisticians. We are grateful to the many colleagues and students who have provided helpful comments, and we hope that they will find this new edition another step forward. In particular, we would like to thank the following colleagues who offered specific comments on the new edition:

Ali Arab *Georgetown University*

Sanjib Basu *Northern Illinois University*

Mary Ellen Bock *Purdue University*

Max Buot *Xavier University*

Jerry J. Chen *Suffolk Community College*

Pinyuen Chen *Syracuse University*

Scott Crawford *University of Wyoming*

Carolyn K. Cuff *Westminster College*

K. L. D. Gunawardena *University of Wisconsin–Oshkosh*

C. Clinton Harshaw *Presbyterian College*

James Helmreich *Marist College*

Ulrich Hoensch *Rocky Mountain College*

Jeff Hovermill *Northern Arizona University*

Debra L. Hydorn *University of Mary Washington*

Monica Jackson *American University*

Tiffany Kolba *Valparaiso University*
Sharon Navard *College of New Jersey*

Ronald C. Neath *Hunter College of CUNY*

Esther M. Pearson *Lasell College*

Thomas Pfaff *Ithaca College*
Kathryn Prewitt *Arizona State University*

Robert L. Sims *George Mason University*

Thomas M. Songer *Portland Community College*

Haiyan Su *Montclair State University*

Anatoliy Swishchuk *University of Calgary*

Frederick C. Tinsley *Colorado College*

Terri Torres *Oregon Institute of Technology*

The professionals at W. H. Freeman and Company, in particular Ruth Baruth, Karen Carson, Katrina Wilhelm, Liam Ferguson, Elizabeth Geller, Vicki Tomaselli, and Lawrence Guerra, have contributed greatly to the success of *IPS*. In addition, we would like to thank Pamela Bruton, Jackie Miller, and Patricia Humphrey for their valuable contributions to the eighth edition. Most of all, we are grateful to the many friends and collaborators whose data and research questions have enabled us to gain a deeper understanding of the science of data. Finally, we would like to acknowledge the contributions of John W. Tukey, whose contributions to data analysis have had such a great influence on us as well as a whole generation of applied statisticians.



W. H. Freeman's new online homework system, **LaunchPad**, offers our quality content curated and organized for easy assignability in a simple but powerful interface. We've taken what we've learned from thousands of instructors and hundreds of thousands of students to create a new generation of W. H. Freeman/Macmillan technology.

Curated Units. Combining a curated collection of videos, homework sets, tutorials, applets, and e-Book content, LaunchPad's interactive units give you a building block to use as is or as a starting point for your own learning units. Thousands of exercises from the text can be assigned as online homework, including many algorithmic exercises. An entire unit's worth of work can be assigned in seconds, drastically reducing the amount of time it takes for you to have your course up and running.

Easily customizable. You can customize the LaunchPad Units by adding quizzes and other activities from our vast wealth of resources. You can also add a discussion board, a dropbox, and RSS feed, with a few clicks. LaunchPad allows you to customize your students' experience as much or as little as you like.

Useful analytics. The gradebook quickly and easily allows you to look up performance metrics for classes, individual students, and individual assignments.

Intuitive interface and design. The student experience is simplified. Students' navigation options and expectations are clearly laid out at all times, ensuring they can never get lost in the system.

Assets integrated into LaunchPad include:

Interactive e-Book. Every LaunchPad e-Book comes with powerful study tools for students, video and multimedia content, and easy customization for instructors. Students can search, highlight, and bookmark, making it easier to study and access key content. And teachers can ensure that their classes get just the book they want to deliver: customize and rearrange chapters, add and share notes and discussions, and link to quizzes, activities, and other resources.



LearningCurve provides students and instructors with powerful adaptive quizzing, a game-like format, direct links to the e-Book, and instant feedback. The quizzing system features questions tailored specifically to the text and adapts to students' responses, providing material at different difficulty levels and topics based on student performance.



SolutionMaster offers an easy-to-use web-based version of the instructor's solutions, allowing instructors to generate a solution file for any set of homework exercises.

New Stepped Tutorials are centered on algorithmically generated quizzing with step-by-step feedback to help students work their way toward the correct solution. These new exercise tutorials (two to three per chapter) are easily assignable and assessable.

Statistical Video Series consists of StatClips, StatClips Examples, and Statistically Speaking “Snapshots.” View animated lecture videos, whiteboard lessons, and documentary-style footage that illustrate key statistical concepts and help students visualize statistics in real-world scenarios.

New Video Technology Manuals available for TI-83/84 calculators, Minitab, Excel, JMP, SPSS, R, Rcmdr, and CrunchIT![®] provide brief instructions for using specific statistical software.

Updated StatTutor Tutorials offer multimedia tutorials that explore important concepts and procedures in a presentation that combines video, audio, and interactive features. The newly revised format includes built-in, assignable assessments and a bright new interface.



Updated Statistical Applets give students hands-on opportunities to familiarize themselves with important statistical concepts and procedures, in an interactive setting that allows them to manipulate variables and see the results graphically. Icons in the textbook indicate when an applet is available for the material being covered.

CrunchIT![®] is a web-based statistical program that allows users to perform all the statistical operations and graphing needed for an introductory statistics course and more. It saves users time by automatically loading data from *IPS* 8e, and it provides the flexibility to edit and import additional data.

Stats@Work Simulations put students in the role of the statistical consultant, helping them better understand statistics interactively within the context of real-life scenarios.

EESEE Case Studies (Electronic Encyclopedia of Statistical Examples and Exercises), developed by The Ohio State University Statistics Department, teach students to apply their statistical skills by exploring actual case studies using real data.

Data files are available in ASCII, Excel, TI, Minitab, SPSS (an IBM Company),* and JMP formats.

Student Solutions Manual provides solutions to the odd-numbered exercises in the text. Available electronically within LaunchPad, as well as in print form.

Interactive Table Reader allows students to use statistical tables interactively to seek the information they need.

Instructor’s Guide with Full Solutions includes teaching suggestions, chapter comments, and detailed solutions to all exercises. Available electronically within LaunchPad, as well as on the IRCD and in print form.

Test Bank offers hundreds of multiple-choice questions. Also available on CD-ROM (for Windows and Mac), where questions can be downloaded, edited, and resequenced to suit each instructor’s needs.

Lecture PowerPoint Slides offer a detailed lecture presentation of statistical concepts covered in each chapter of *IPS*.

*SPSS was acquired by IBM in October 2009.

Additional Resources Available with IPS 8e

Companion Website www.whfreeman.com/ips8e This open-access website includes statistical applets, data files, supplementary exercises, and self-quizzes. The website also offers four optional companion chapters covering logistic regression, nonparametric tests, bootstrap methods and permutation tests, and statistics for quality control and capability.

Instructor access to the Companion Website requires user registration as an instructor and features all of the open-access student web materials, plus:

- Instructor version of **EESSE** with solutions to the exercises in the student version.
- **PowerPoint Slides** containing all textbook figures and tables.
- **Lecture PowerPoint Slides**

Special Software Packages Student versions of JMP and Minitab are available for packaging with the text. Contact your W. H. Freeman representative for information or visit www.whfreeman.com.

Enhanced Instructor's Resource CD-ROM, ISBN: 1-4641-3360-3 Allows instructors to **search** and **export** (by key term or chapter) all the resources available on the student companion website plus the following:

- All text images and tables
- Instructor's Guide with Full Solutions
- PowerPoint files and lecture slides
- Test Bank files

Course Management Systems W. H. Freeman and Company provides courses for Blackboard, Angel, Desire2Learn, Canvas, Moodle, and Sakai course management systems. These are completely integrated solutions that you can easily customize and adapt to meet your teaching goals and course objectives. Visit macmillanhighered.com/Catalog/other/Coursepack for more information.

iClicker **iClicker** is a two-way radio-frequency classroom response solution developed by educators for educators. Each step of i-clicker's development has been informed by teaching and learning. To learn more about packaging i-clicker with this textbook, please contact your local sales rep or visit www.iclicker.com.

Statistics is the science of collecting, organizing, and interpreting numerical facts, which we call *data*. We are bombarded by data in our everyday lives. The news mentions movie box-office sales, the latest poll of the president's popularity, and the average high temperature for today's date. Advertisements claim that data show the superiority of the advertiser's product. All sides in public debates about economics, education, and social policy argue from data. A knowledge of statistics helps separate sense from nonsense in this flood of data.

The study and collection of data are also important in the work of many professions, so training in the science of statistics is valuable preparation for a variety of careers. Each month, for example, government statistical offices release the latest numerical information on unemployment and inflation. Economists and financial advisers, as well as policy makers in government and business, study these data in order to make informed decisions. Doctors must understand the origin and trustworthiness of the data that appear in medical journals. Politicians rely on data from polls of public opinion. Business decisions are based on market research data that reveal consumer tastes and preferences. Engineers gather data on the quality and reliability of manufactured products. Most areas of academic study make use of numbers and, therefore, also make use of the methods of statistics. This means it is extremely likely that your undergraduate research projects will involve, at some level, the use of statistics.

Learning from Data

The goal of statistics is to learn from data. To learn, we often perform calculations or make graphs based on a set of numbers. But to learn from data, we must do more than calculate and plot, because data are not just numbers; they are numbers that have some context that helps us learn from them.

Two-thirds of Americans are overweight or obese according to the Centers for Disease Control and Prevention (CDC) website (www.cdc.gov/nchs/nhanes.htm). What does it mean to be obese or to be overweight? To answer this question we need to talk about body mass index (BMI). Your weight in kilograms divided by the square of your height in meters is your BMI. A man who is 6 feet tall (1.83 meters) and weighs 180 pounds (81.65 kilograms) will have a BMI of $81.65/(1.83)^2 = 24.4 \text{ kg/m}^2$. How do we interpret this number? According to the CDC, a person is classified as overweight if his or her BMI is between 25 and 29 kg/m^2 and as obese if his or her BMI is 30 kg/m^2 or more. Therefore, two-thirds of Americans have a BMI of 25 kg/m^2 or more. The man who weighs 180 pounds and is 6 feet tall is not overweight or obese, but if he gains 5 pounds, his BMI would increase to 25.1, and he would be classified as overweight.

When you do statistical problems, even straightforward textbook problems, don't just graph or calculate. Think about the context and state your conclusions in the specific setting of the problem. As you are learning how to do statistical calculations and graphs, remember that the goal of statistics is not calculation for its own sake but gaining understanding from numbers. The calculations and graphs can be automated by a calculator or software,

but you must supply the understanding. This book presents only the most common specific procedures for statistical analysis. A thorough grasp of the principles of statistics will enable you to quickly learn more advanced methods as needed. On the other hand, a fancy computer analysis carried out without attention to basic principles will often produce elaborate nonsense. As you read, seek to understand the principles as well as the necessary details of methods and recipes.

The Rise of Statistics

Historically, the ideas and methods of statistics developed gradually as society grew interested in collecting and using data for a variety of applications. The earliest origins of statistics lie in the desire of rulers to count the number of inhabitants or measure the value of taxable land in their domains. As the physical sciences developed in the seventeenth and eighteenth centuries, the importance of careful measurements of weights, distances, and other physical quantities grew. Astronomers and surveyors striving for exactness had to deal with variation in their measurements. Many measurements should be better than a single measurement, even though they vary among themselves. How can we best combine many varying observations? Statistical methods that are still important were invented in order to analyze scientific measurements.

By the nineteenth century, the agricultural, life, and behavioral sciences also began to rely on data to answer fundamental questions. How are the heights of parents and children related? Does a new variety of wheat produce higher yields than the old, and under what conditions of rainfall and fertilizer? Can a person's mental ability and behavior be measured just as we measure height and reaction time? Effective methods for dealing with such questions developed slowly and with much debate.

As methods for producing and understanding data grew in number and sophistication, the new discipline of statistics took shape in the twentieth century. Ideas and techniques that originated in the collection of government data, in the study of astronomical or biological measurements, and in the attempt to understand heredity or intelligence came together to form a unified "science of data." That science of data—statistics—is the topic of this text.

The Organization of This Book

Part I of this book, called simply "Looking at Data," concerns data analysis and data production. The first two chapters deal with statistical methods for organizing and describing data. These chapters progress from simpler to more complex data. Chapter 1 examines data on a single variable, Chapter 2 is devoted to relationships among two or more variables. You will learn both how to examine data produced by others and how to organize and summarize your own data. These summaries will first be graphical, then numerical, and then, when appropriate, in the form of a mathematical model that gives a compact description of the overall pattern of the data. Chapter 3 outlines arrangements (called designs) for producing data that answer specific questions. The principles presented in this chapter will help you to design proper samples and experiments for your research projects and to evaluate other such investigations in your field of study.

Part II, consisting of Chapters 4 to 8, introduces statistical inference—formal methods for drawing conclusions from properly produced data. Statistical inference uses the language of probability to describe how reliable its conclusions are, so some basic facts about probability are needed to understand inference. Probability is the subject of Chapters 4 and 5. Chapter 6, perhaps the most important chapter in the text, introduces the reasoning of statistical inference. Effective inference is based on good procedures for producing data (Chapter 3), careful examination of the data (Chapters 1 and 2), and an understanding of the nature of statistical inference as discussed in Chapter 6. Chapters 7 and 8 describe some of the most common specific methods of inference, for drawing conclusions about means and proportions from one and two samples.

The five shorter chapters in Part III introduce somewhat more advanced methods of inference, dealing with relations in categorical data, regression and correlation, and analysis of variance. Four supplementary chapters, available from the text website, present additional statistical topics.

What Lies Ahead

Introduction to the Practice of Statistics is full of data from many different areas of life and study. Many exercises ask you to express briefly some understanding gained from the data. In practice, you would know much more about the background of the data you work with and about the questions you hope the data will answer. No textbook can be fully realistic. But it is important to form the habit of asking, “What do the data tell me?” rather than just concentrating on making graphs and doing calculations.

You should have some help in automating many of the graphs and calculations. You should certainly have a calculator with basic statistical functions. Look for keywords such as “two-variable statistics” or “regression” when you shop for a calculator. More advanced (and more expensive) calculators will do much more, including some statistical graphs. You may be asked to use software as well. There are many kinds of statistical software, from spreadsheets to large programs for advanced users of statistics. The kind of computing available to learners varies a great deal from place to place—but the big ideas of statistics don’t depend on any particular level of access to computing.

Because graphing and calculating are automated in statistical practice, the most important assets you can gain from the study of statistics are an understanding of the big ideas and the beginnings of good judgment in working with data. Ideas and judgment can’t (at least yet) be automated. They guide you in telling the computer what to do and in interpreting its output. This book tries to explain the most important ideas of statistics, not just teach methods. Some examples of big ideas that you will meet are “always plot your data,” “randomized comparative experiments,” and “statistical significance.”

You learn statistics by doing statistical problems. “Practice, practice, practice.” Be prepared to work problems. The basic principle of learning is persistence. Being organized and persistent is more helpful in reading this book than knowing lots of math. The main ideas of statistics, like the main ideas of any important subject, took a long time to discover and take some time to master. The gain will be worth the pain.

ABOUT THE AUTHORS

David S. Moore is Shanti S. Gupta Distinguished Professor of Statistics, Emeritus, at Purdue University and was 1998 president of the American Statistical Association. He received his AB from Princeton and his PhD from Cornell, both in mathematics. He has written many research papers in statistical theory and served on the editorial boards of several major journals. Professor Moore is an elected fellow of the American Statistical Association and of the Institute of Mathematical Statistics and an elected member of the International Statistical Institute. He has served as program director for statistics and probability at the National Science Foundation.

In recent years, Professor Moore has devoted his attention to the teaching of statistics. He was the content developer for the Annenberg/Corporation for Public Broadcasting college-level telecourse *Against All Odds: Inside Statistics* and for the series of video modules *Statistics: Decisions through Data*, intended to aid the teaching of statistics in schools. He is the author of influential articles on statistics education and of several leading texts. Professor Moore has served as president of the International Association for Statistical Education and has received the Mathematical Association of America's national award for distinguished college or university teaching of mathematics.

George P. McCabe is Associate Dean for Academic Affairs in the College of Science and Professor of Statistics at Purdue University. In 1966, he received a BS degree in mathematics from Providence College and in 1970 a PhD in mathematical statistics from Columbia University. His entire professional career has been spent at Purdue, with sabbaticals at Princeton University, the Commonwealth Scientific and Industrial Research Organization (CSIRO) in Melbourne (Australia), the University of Berne (Switzerland), the National Institute of Standards and Technology (NIST) in Boulder, Colorado, and the National University of Ireland in Galway. Professor McCabe is an elected fellow of the American Association for the Advancement of Science and of the American Statistical Association; he was 1998 Chair of its section on Statistical Consulting. In 2008–2010, he served on the Institute of Medicine Committee on Nutrition Standards for the National School Lunch and Breakfast Programs. He has served on the editorial boards of several statistics journals. He has consulted with many major corporations and has testified as an expert witness on the use of statistics in several cases.

Professor McCabe's research interests have focused on applications of statistics. Much of his recent work has focused on problems in nutrition, including nutrient requirements, calcium metabolism, and bone health. He is the author or coauthor of over 170 publications in many different journals.

Bruce A. Craig is Professor of Statistics and Director of the Statistical Consulting Service at Purdue University. He received his BS in mathematics and economics from Washington University in St. Louis and his PhD in statistics from the University of Wisconsin–Madison. He is an elected fellow of the American Statistical Association and was chair of its section on Statistical Consulting in 2009. He is also an active member of the Eastern North American Region of the International Biometrics Society and was elected by the voting membership to the Regional Committee between 2003 and 2006. Professor Craig has served on the editorial board of several statistical journals and has been a member of several data and safety monitoring boards, including Purdue’s institutional review board.

Professor Craig’s research interests focus on the development of novel statistical methodology to address research questions in the life sciences. Areas of current interest are protein structure determination, diagnostic testing, and animal abundance estimation. In 2005, he was named Purdue University Faculty Scholar.

DATA TABLE INDEX

TABLE 1.1	IQ test scores for 60 randomly chosen fifth-grade students	16
TABLE 1.2	Service times (seconds) for calls to a customer service center	19
TABLE 1.3	Educational data for 78 seventh-grade students	29
TABLE 2.1	World record times for the 10,000-meter run	102
TABLE 2.2	Four data sets for exploring correlation and regression	125
TABLE 2.3	Two measures of glucose level in diabetics	130
TABLE 2.4	Dwelling permits, sales, and production for 21 European countries	159
TABLE 2.5	Fruit and vegetable consumption and smoking	164
TABLE 7.1	Monthly rates of return on a portfolio (%)	425
TABLE 7.2	Aggressive behaviors of dementia patients	429
TABLE 7.3	Length (in seconds) of audio files sampled from an iPod	437
TABLE 7.4	DRP scores for third-graders	452
TABLE 7.5	Seated systolic blood pressure (mm Hg)	463
TABLE 10.1	In-state tuition and fees (in dollars) for 33 public universities	602
TABLE 10.2	Sales price and assessed value (in \$ thousands) of 30 homes in a midwestern city	604
TABLE 10.3	Annual number of tornadoes in the United States between 1953 and 2012	605
TABLE 10.4	Watershed area (km ²), percent forest, and index of biotic integrity	606
TABLE 12.1	Age at death for North American women writers	686
TABLE 13.1	Safety behaviors of abused women	713
TABLE 13.2	Iron content (mg/100 g) of food cooked in different pots	716
TABLE 13.3	Tool diameter data	716
TABLE 16.1	Degree of Reading Power scores for third-graders	16-44
TABLE 16.2	Aggressive behaviors of dementia patients	16-50
TABLE 16.3	Serum retinol levels in two groups of children	16-55
TABLE 17.1	Twenty control chart samples of water resistance	17-10
TABLE 17.2	Control chart constants	17-14

TABLE 17.3	Twenty samples of size 3, with \bar{x} and s	17-19
TABLE 17.4	Three sets of \bar{x} 's from 20 samples of size 4	17-20
TABLE 17.5	Twenty samples of size 4, with \bar{x} and s	17-21
TABLE 17.6	\bar{x} and s for 24 samples of elastomer viscosity	17-27
TABLE 17.7	\bar{x} and s for 24 samples of label placement	17-36
TABLE 17.8	\bar{x} and s for 24 samples of label placement	17-37
TABLE 17.9	Hospital losses for 15 samples of DRG 209 patients	17-38
TABLE 17.10	Daily calibration samples for a Lunar bone densitometer	17-39
TABLE 17.11	\bar{x} and s for samples of bore diameter	17-40
TABLE 17.12	Fifty control chart samples of call center response times	17-50
TABLE 17.13	Proportions of workers absent during four weeks	17-56
TABLE 17.14	\bar{x} and s for samples of film thickness	17-59

BEYOND THE BASICS INDEX

CHAPTER 1	Density estimation	71
CHAPTER 2	Scatterplot smoothers	96
CHAPTER 2	Data mining	135
CHAPTER 3	Capture-recapture sampling	214
CHAPTER 4	More laws of large numbers	270
CHAPTER 5	Weibull distributions	315
CHAPTER 6	The bootstrap	367
CHAPTER 8	The plus four confidence interval for a single proportion	493
CHAPTER 8	The plus four confidence interval for a difference in proportions	493
CHAPTER 8	Relative risk	520
CHAPTER 9	Meta-analysis	548
CHAPTER 10	Nonlinear regression	582
CHAPTER 11	Multiple logistic regression	631
CHAPTER 16	The bootstrap for a scatterplot smoother	16-20



Looking at Data—Distributions

1

Introduction

CHAPTER

Statistics is the science of learning from data. Data are numerical or qualitative descriptions of the objects that we want to study. In this chapter, we will master the art of examining data.

We begin in Section 1.1 with some basic ideas about data. We will learn about the different types of data that are collected and how data sets are organized.

Section 1.2 starts our process of learning from data by looking at graphs. These visual displays give us a picture of the overall patterns in a set of data. We have excellent software tools that help us make these graphs. However, it takes a little experience and a lot of judgment to study the graphs carefully and to explain what they tell us about our data.

Section 1.3 continues our process of learning from data by computing numerical summaries. These sets of numbers describe key characteristics of the patterns that we saw in our graphical summaries.

A final section in this chapter helps us make the transition from data summaries to statistical models. We learn about using density curves to describe a set of data. The Normal distributions are also introduced in this section. These distributions can be used to describe many sets of data that we will encounter. They also play a fundamental role in the methods that we will use to draw conclusions from many sets of data.

- 1.1 Data
- 1.2 Displaying Distributions with Graphs
- 1.3 Describing Distributions with Numbers
- 1.4 Density Curves and Normal Distributions

1.1 Data

When you complete this section, you will be able to

- Give examples of cases in a data set.
- Identify the variables in a data set.
- Demonstrate how a label can be used as a variable in a data set.
- Identify the values of a variable.
- Classify variables as categorical or quantitative.
- Describe the key characteristics of a set of data.
- Explain how a rate is the result of adjusting one variable to create another.

A statistical analysis starts with a set of data. We construct a set of data by first deciding what *cases*, or units, we want to study. For each case, we record information about characteristics that we call *variables*.

CASES, LABELS, VARIABLES, AND VALUES

Cases are the objects described by a set of data. Cases may be customers, companies, subjects in a study, units in an experiment, or other objects.

A **label** is a special variable used in some data sets to distinguish the different cases.

A **variable** is a characteristic of a case.

Different cases can have different **values** of the variables.

EXAMPLE

1.1 Over 12 billion sold. Apple’s music-related products and services generated \$1.8 billion in the third quarter of 2012. Since Apple started marketing iTunes in 2003, they have sold over 12 billion songs. Let’s take a look at this remarkable product. Figure 1.1 is part of an iTunes playlist named IPS. The six songs shown are *cases*. They are numbered from 1 to 6 in the first column. These numbers are the *labels* that distinguish the six songs. The following five columns give name (of the song), time (the length of time it takes to play the song), artist, album, and genre.

Some variables, like the name of a song and the artist simply place cases into categories. Others, like the length of a song, take numerical values for which we can do arithmetic. It makes sense to give an average length of time for a collection of songs, but it does not make sense to give an “average” album. We can, however, count the numbers of songs on different albums, and we can do arithmetic with these counts.



FIGURE 1.1 Part of an iTunes playlist, for Example 1.1.

CATEGORICAL AND QUANTITATIVE VARIABLES

A **categorical variable** places a case into one of several groups or categories.

A **quantitative variable** takes numerical values for which arithmetic operations such as adding and averaging make sense.

The **distribution** of a variable tells us what values it takes and how often it takes these values.

EXAMPLE

1.2 Categorical and quantitative variables in iTunes playlist. The IPS iTunes playlist contains five variables. These are the name, time, artist, album, and genre. The time is a quantitative variable. Name, artist, album, and genre are categorical variables.



An appropriate label for your cases should be chosen carefully. In our iTunes example, a natural choice of a label would be the name of the song. However, if you have more than one artist performing the same song, or the same artist performing the same song on different albums, then the name of the song would not uniquely label each of the songs in your playlist.



A quantitative variable such as the time in the iTunes playlist requires some special attention before we can do arithmetic with its values. The first song in the playlist has time equal to 3:32—that is, 3 minutes and 32 seconds. To do arithmetic with this variable, we should first convert all the values so that they have a single unit. We could convert to seconds; 3 minutes is 180 seconds, so the total time is $180 + 32$, or 212 seconds. An alternative would be to convert to minutes; 32 seconds is 0.533 minute, so time written in this way is 3.533 minutes.

USE YOUR KNOWLEDGE

1.1 Time in the iTunes playlist. In the iTunes playlist, do you prefer to convert the time to seconds or minutes? Give a reason for your answer.

units of measurement

We use the term **units of measurement** to refer to the seconds or minutes that tell us how the variable time is measured. If we were measuring heights of children, we might choose to use either inches or centimeters. The units of measurement are an important part of the description of a quantitative variable.

Key characteristics of a data set

In practice, any set of data is accompanied by background information that helps us understand the data. When you plan a statistical study or explore data from someone else's work, ask yourself the following questions:

1. **Who?** What **cases** do the data describe? **How many** cases does the data set contain?
2. **What?** How many **variables** do the data contain? What are the **exact definitions** of these variables? What are the units of measurement for each quantitative variable?
3. **Why? What purpose** do the data have? Do we hope to answer some specific questions? Do we want to draw conclusions about cases other than the ones we actually have data for? Are the variables that are recorded suitable for the intended purpose?

EXAMPLE



1.3 Data for students in a statistics class. Figure 1.2 shows part of a data set for students enrolled in an introductory statistics class. Each row gives the data on one student. The values for the different variables are in the columns. This data set has eight variables. ID is a label for each student. Exam1, Exam2, Homework, Final, and Project give the points earned, out of a total of 100 possible, for each of these course requirements. Final grades are based on a possible 200 points for each exam and the Final, 300 points for Homework, and 100 points for Project. TotalPoints is the variable that gives the composite score. It is computed by adding 2 times Exam1, Exam2, and Final, 3 times Homework, and 1 times Project. Grade is the grade earned in the course. This instructor used cutoffs of 900, 800, 700, etc. for the letter grades.

USE YOUR KNOWLEDGE

- 1.2 **Who, what, and why for the statistics class data.** Answer the who, what, and why questions for the statistics class data set.
- 1.3 **Read the spreadsheet.** Refer to Figure 1.2. Give the values of the variables Exam1, Exam2, and Final for the student with ID equal to 104.

FIGURE 1.2 Spreadsheet for Example 1.3.

	A	B	C	D	E	F	G	H
1	ID	Exam1	Exam2	Homework	Final	Project	TotalPoints	Grade
2	101	89	94	88	87	95	899	B
3	102	78	84	90	89	94	866	B
4	103	71	80	75	79	95	780	C
5	104	95	98	97	96	93	962	A
6	105	79	88	85	88	96	861	B

1.4 Calculate the grade. A student whose data do not appear on the spreadsheet scored 83 on Exam1, 82 on Exam2, 77 for Homework, 90 on the Final, and 80 on the Project. Find TotalPoints for this student and give the grade earned.

spreadsheet

The display in Figure 1.2 is from an Excel **spreadsheet**. Spreadsheets are very useful for doing the kind of simple computations that you did in Exercise 1.4. You can type in a formula and have the same computation performed for each row.

Note that the names we have chosen for the variables in our spreadsheet do not have spaces. For example, we could have used the name “Exam 1” for the first-exam score rather than Exam1. In some statistical software packages, however, spaces are not allowed in variable names. *For this reason, when creating spreadsheets for eventual use with statistical software, it is best to avoid spaces in variable names.* Another convention is to use an underscore (_) where you would normally use a space. For our data set, we could use Exam_1, Exam_2, and Final_Exam.



EXAMPLE

1.4 Cases and variables for the statistics class data. The data set in Figure 1.2 was constructed to keep track of the grades for students in an introductory statistics course. The cases are the students in the class. There are eight variables in this data set. These include a label for each student and scores for the various course requirements. There are no units for ID and grade. The other variables all have “points” as the unit.

EXAMPLE

1.5 Statistics class data for a different purpose. Suppose that the data for the students in the introductory statistics class were also to be used to study relationships between student characteristics and success in the course. For this purpose, we might want to use a data set like the spreadsheet in Figure 1.3.

FIGURE 1.3 Spreadsheet for Example 1.5.

	A	B	C	D	E	F
1	ID	TotalPoints	Grade	Gender	PrevStat	Year
2	101	899	B	F	Yes	4
3	102	866	B	M	Yes	3
4	103	780	C	M	No	3
5	104	962	A	M	No	1
6	105	861	B	F	No	4

Here, we have decided to focus on the TotalPoints and Grade as the outcomes of interest. Other variables of interest have been included: Gender, PrevStat (whether or not the student has taken a statistics course previously), and Year (student classification as first, second, third, or fourth year). ID is a categorical variable, TotalPoints is a quantitative variable, and the remaining variables are all categorical.

In our example, the possible values for the grade variable are A, B, C, D, and F. When computing grade point averages, many colleges and universities translate these letter grades into numbers using $A = 4$, $B = 3$, $C = 2$, $D = 1$, and $F = 0$. The transformed variable with numeric values is considered to be quantitative because we can average the numerical values across different courses to obtain a grade point average.

Sometimes, experts argue about numerical scales such as this. They ask whether or not the difference between an A and a B is the same as the difference between a D and an F. Similarly, many questionnaires ask people to respond on a 1 to 5 scale with 1 representing strongly agree, 2 representing agree, etc. Again, we could ask whether or not the five possible values for this scale are equally spaced in some sense. From a practical point of view, however, the averages that can be computed when we convert categorical scales such as these to numerical values frequently provide a very useful way to summarize data.

USE YOUR KNOWLEDGE

1.5 Apartment rentals. A data set lists apartments available for students to rent. Information provided includes the monthly rent, whether or not cable is included free of charge, whether or not pets are allowed, the number of bedrooms, and the distance to the campus. Describe the cases in the data set, give the number of variables, and specify whether each variable is categorical or quantitative.

Often the variables in a statistical study are easy to understand: height in centimeters, study time in minutes, and so on. But each area of work also has its own special variables. A psychologist uses the Minnesota Multiphasic Personality Inventory (MMPI), and a physical fitness expert measures “VO2 max,” the volume of oxygen consumed per minute while exercising at your maximum capacity. Both of these variables are measured with special **instruments**. VO2 max is measured by exercising while breathing into a mouthpiece connected to an apparatus that measures oxygen consumed. Scores on the MMPI are based on a long questionnaire, which is also an instrument.

Part of mastering your field of work is learning what variables are important and how they are best measured. Because details of particular measurements usually require knowledge of the particular field of study, we will say little about them.

Be sure that each variable really does measure what you want it to. A poor choice of variables can lead to misleading conclusions. Often, for example, the **rate** at which something occurs is a more meaningful measure than a simple count of occurrences.

instrument



rate

EXAMPLE

1.6 Comparing colleges based on graduates. Think about comparing colleges based on the numbers of graduates. This view tells you something about the relative sizes of different colleges. However, if you are interested in how well colleges succeed at graduating students whom they admit, it would be better to use a rate. For example, you can find data on the Internet on the six-year graduation rates of different colleges. These rates are computed by examining the progress of first-year students who enroll in a given year. Suppose that at College A there were 1000 first-year students in a particular year, and 800 graduated within six years. The graduation rate is

$$\frac{800}{1000} = 0.80$$

or 80%. College B has 2000 students who entered in the same year, and 1200 graduated within six years. The graduation rate is

$$\frac{1200}{2000} = 0.60$$

or 60%. How do we compare these two colleges? College B has more graduates, but College A has a better graduation rate.

USE YOUR KNOWLEDGE

1.6 How should you express the change? Between the first exam and the second exam in your statistics course you increased the amount of time that you spent working exercises. Which of the following three ways would you choose to express the results of your increased work: (a) give the grades on the two exams, (b) give the ratio of the grade on the second exam divided by the grade on the first exam, or (c) take the difference between the grade on the second exam and the grade on the first exam, and express this as a percent of the grade on the first exam. Give reasons for your answer.

1.7 Which variable would you choose. Refer to Example 1-6, on colleges and their graduates.

(a) Give a setting where you would prefer to evaluate the colleges based on the numbers of graduates. Give a reason for your choice.

(b) Give a setting where you would prefer to evaluate the colleges based on the graduation rates. Give a reason for your choice.

In Example 1.6, when we computed the graduation rate, we used the total number of students to adjust the number of graduates. We constructed a new variable by dividing the number of graduates by the total number of students. Computing a rate is just one of several ways of **adjusting one variable to create another**. We often divide one variable by another to compute a more meaningful variable to study. Example 1.20 (page 22) is another type of adjustment.

adjusting one variable
to create another



Exercises 1.6 and 1.7 illustrate an important point about presenting the results of your statistical calculations. *Always consider how to best communicate your results to a general audience.* For example, the numbers produced by your calculator or by statistical software frequently contain more digits than are needed. Be sure that you do not include extra information generated by software that will distract from a clear explanation of what you have found.

SECTION 1.1 Summary

A data set contains information on a number of **cases**. Cases may be customers, companies, subjects in a study, units in an experiment, or other objects.

For each case, the data give values for one or more **variables**. A variable describes some characteristic of a case, such as a person's height, gender, or salary. Variables can have different **values** for different cases.

A **label** is a special variable used to identify cases in a data set.

Some variables are **categorical** and others are **quantitative**. A categorical variable places each individual into a category, such as male or female. A quantitative variable has numerical values that measure some characteristic of each case, such as height in centimeters or annual salary in dollars.

The **key characteristics** of a data set answer the questions Who?, What?, and Why?

SECTION 1.1 Exercises

For Exercise 1.1, see page 3; for Exercises 1.2 to 1.4, see pages 4–5; for Exercise 1.5, see page 6; and for Exercises 1.6 and 1.7, see page 7.

1.8 Summer jobs. You are collecting information about summer jobs that are available for college students in your area. Describe a data set that you could use to organize the information that you collect.

- What are the cases?
- Identify the variables and their possible values.
- Classify each variable as categorical or quantitative. Be sure to include at least one of each.
- Use a label and explain how you chose it.
- Summarize the key characteristics of your data set.

1.9 Employee application data. The personnel department keeps records on all employees in a company. Here is the information that they keep in one of their data files: employee identification number, last name, first name, middle initial, department, number of years with the company, salary, education (coded as high school, some college, or college degree), and age.

- What are the cases for this data set?
- Describe each type of information as a label, a quantitative variable, or a categorical variable.

(c) Set up a spreadsheet that could be used to record the data. Give appropriate column headings and five sample cases.

1.10 How would you rank cities? Various organizations rank cities and produce lists of the 10 or the 100 best based on various measures. Create a list of criteria that you would use to rank cities. Include at least eight variables and give reasons for your choices. Say whether each variable is quantitative or categorical.

1.11 Survey of students. A survey of students in an introductory statistics class asked the following questions: (1) age; (2) do you like to sing? (Yes, No); (3) can you play a musical instrument (not at all, a little, pretty well); (4) how much did you spend on food last week? (5) height.

- Classify each of these variables as categorical or quantitative and give reasons for your answers.
- For each variable give the possible values.

1.12 What questions would you ask? Refer to the previous exercise. Make up your own survey questions with at least six questions. Include at least two categorical variables and at least two quantitative variables. Tell which variables are categorical and which are quantitative. Give reasons for your answers. For each variable give the possible values.

1.13 How would you rate colleges? Popular magazines rank colleges and universities on their “academic quality” in serving undergraduate students. Describe five variables that you would like to see measured for each college if you were choosing where to study. Give reasons for each of your choices.

1.14 Attending college in your state or in another state. The U.S. Census Bureau collects a large amount of information concerning higher education.¹ For example, the bureau provides a table that includes the following variables: state, number of students from the state who attend college, number of students who attend college in their home state.

- (a) What are the cases for this set of data?
 (b) Is there a label variable? If yes, what is it?

- (c) Identify each variable as categorical or quantitative.
 (d) Explain how you might use each of the quantitative variables to explain something about the states.
 (e) Consider a variable computed as the number of students in each state who attend college in the state divided by the total number of students from the state who attend college. Explain how you would use this variable explain something about the states.

1.15 Alcohol-impaired driving fatalities. A report on drunk-driving fatalities in the United States gives the number of alcohol-impaired driving fatalities for each state.² Discuss at least three different ways that these numbers could be converted to rates. Give the advantages and disadvantages of each.

1.2 Displaying Distributions with Graphs

When you complete this section, you will be able to

- Analyze the distribution of a categorical variable using a bar graph.
- Analyze the distribution of a categorical variable using a pie chart.
- Analyze the distribution of a quantitative variable using a stemplot.
- Analyze the distribution of a quantitative variable using a histogram.
- Examine the distribution of a quantitative variable with respect to the overall pattern of the data and deviations from that pattern.
- Identify the shape, center, and spread of the distribution of a quantitative variable.
- Identify and describe any outliers in the distribution of a quantitative variable.
- Use a time plot to describe the distribution of a quantitative variable that is measured over time.

exploratory data analysis

Statistical tools and ideas help us examine data to describe their main features. This examination is called **exploratory data analysis**. Like an explorer crossing unknown lands, we want first to simply describe what we see. Here are two basic strategies that help us organize our exploration of a set of data:

- Begin by examining each variable by itself. Then move on to study the relationships among the variables.
- Begin with a graph or graphs. Then add numerical summaries of specific aspects of the data.