时代教育·国外高校优秀教材精选

(英文版)

应用统计分析
——使用Excel

Applied Statistics with Microsoft Excel

(美)杰拉尔德 凯勒(Gerald Keller) 著

附赠光盘

机械工业出版社
CHINA MACHINE PRESS

时代教育·国外高校优秀教材精选

应用统计分析

——使用 Excel

Applied Statistics with Microsoft Excel

（英文版）

（美） Gerald Keller（杰拉尔德　凯勒）著

机械工业出版社

Gerald Keller

Applied Statistics with Microsoft Excel

EISBN: 0-534-37112-4

981-254-435-6

北京市版权局著作权合同登记号：图字 01-2004-0888 号

图书在版编目(CIP)数据

应用统计分析：使用 Excel/(美)凯勒(Keller, G.)著. —北京：机械工业出版社，2004.5
(时代教育·国外高校优秀教材精选)
ISBN 7-111-14321-3

Ⅰ. 应 ... Ⅱ. 凯 ... Ⅲ. 电子表格系统，Excel - 应用 - 统计分析 - 高等学校 - 教材 - 英文 Ⅳ. C812

中国版本图书馆 CIP 数据核字(2004)第 030108 号

机械工业出版社(北京市百万庄大街 22 号 邮政编码 100037)
责任编辑：郑 玫 封面设计：饶 薇 责任印制：施 红

北京铭成印刷有限公司印刷·新华书店北京发行所发行

2004 年 5 月第 1 版第 1 次印刷
787mm × 1092mm 1/16·46.25 印张·1 插页·1151 千字
定价：68.00 元(含 1CD)

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
本社购书热线电话(010)68993821、88379646
封面无防伪标均为盗版

A GUIDE TO STATISTICAL TECHNIQUES

DATA TYPES	PROBLEM OBJECTIVES: Describe a Single Population	Compare Two Populations	Compare Two or More Populations	Analyze the Relationship Between Two Variables	Analyze the Relationship Among Two or More Variables
Interval	Histogram **Section 2.3** Box plot **Section 3.5** Mean, median, and mode **Section 3.2** Range, variance, and standard deviation **Section 3.3** Percentiles and quartiles **Section 3.5** t-test and estimator of a mean **Section 11.2** Chi-squared test and estimator of a variance **Section 11.3**	Equal-variances t-test and estimator of difference between two means (independent samples) **Section 12.2** Unequal-variances t-test and estimator of difference between two means (independent samples) **Section 12.2** t-test and estimator of mean difference (matched pairs) **Section 12.4** F-test and estimator of ratio of two variances **Section 12.5** Wilcoxon rank sum test **Section 16.2** Wilcoxon signed rank sum test **Section 16.3**	Analysis of variance: Independent samples, single-factor **Section 14.2** Analysis of variance: Independent samples, two-factor **Section 14.5** Analysis of variance: Randomized blocks **Section 14.4** Kruskal-Wallis test **Section 16.4** Friedman test **Section 16.5** Tukey's multiple comparison method **Section 14.6** LSD multiple comparison method **Section 14.6** Bartlett's test **Section 14.7**	Scatter diagram **Section 2.6** Covariance **Section 3.6** Correlation **Section 3.6** t-test of correlation **Section 17.7** Simple linear regression **Chapter 17** Spearman rank correlation test **Section 17.7**	Multiple regression **Chapter 18**
Nominal	Pie chart **Section 2.4** Bar chart **Section 2.4** Line chart **Section 2.5** z-test and estimator of a proportion **Section 11.4** Chi-squared goodness-of-fit test **Section 15.2**	z-test and estimator of difference between two proportions **Section 12.6** Chi-squared test of contingency table **Section 15.3**	Chi-squared test of contingency table **Section 15.3**	Chi-squared test of contingency table **Section 15.3**	Not covered
Ordinal	Box plot **Section 3.5** Median **Section 3.2** Percentiles and Quartiles **Section 3.5**	Wilcoxon rank sum test **Section 16.2** Sign test **Section 16.3**	Kruskal-Wallis test **Section 16.4** Friedman test **Section 16.5**	Spearman rank correlation test **Section 17.7**	Not covered

国外高校优秀教材审定委员会

出版说明

随着我国加入 WTO，国际间的竞争越来越激烈，而国际间的竞争实际上也就是人才的竞争、教育的竞争。为了加快培养具有国际竞争力的高水平技术人才，加快我国教育改革的步伐，国家教育部近来出台了一系列倡导高校开展双语教学、引进原版教材的政策。以此为契机，机械工业出版社陆续推出了一系列国外影印版教材，其内容涉及高等学校公共基础课，以及机、电、信息领域的专业基础课和专业课。

引进国外优秀原版教材，在有条件的学校推动开展英语授课或双语教学，自然也引进了先进的教学思想和教学方法，这对提高我国自编教材的水平，加强学生的英语实际应用能力，使我国的高等教育尽快与国际接轨，必将起到积极的推动作用。

为了做好教材的引进工作，机械工业出版社特别成立了由著名专家组成的国外高校优秀教材审定委员会。这些专家对实施双语教学做了深入细致的调查研究，对引进原版教材提出了许多建设性意见，并慎重地对每一本将要引进的原版教材一审再审，精选再精选，确认教材本身的质量水平，以及权威性和先进性，以期所引进的原版教材能适应我国学生的外语水平和学习特点。在引进工作中，审定委员会还结合我国高校教学课程体系的设置和要求，对原版教材的教学思想和方法的先进性、科学性严格把关。同时尽量考虑原版教材的系统性和经济性。

这套教材出版后，我们将根据各高校的双语教学计划，举办原版教材的教师培训，及时地将其推荐给各高校选用。希望高校师生在使用教材后及时反馈意见和建议，使我们更好地为教学改革服务。

机械工业出版社

序

本书的作者有多年概率统计学、管理学和运筹管理学的教学经验，编写了多本畅销教材。在本书中，作者继续体现了他的“三步式”解决问题的方法，第一步是 Indentify，包括试验设计、收集数据和选择模型；第二步是 Compute，即用 Excel 做计算；第三步是 Interpret，就是分析、解释计算的结果。

本书使用较少的概率知识，从各个应用层面，通过丰富的案例分析和读者自己动手的应用实例，讲解了应用统计的基本内容。本书系统地阐述了如何正确收集数据资料，如何使用 Microsoft Excel 软件进行统计分析，应如何从中得到有意义的统计结论。使用此书不需要微积分基础，只要具有高中的数学水平就可以通览全书。

此书有以下几个特点：

第一，本书更多的用形象思维与直观判断引进统计概念和方法。

第二，本书层次清楚，每段都有主要的概念和公式的小结和大量的练习题，应用题的图标分明，利于复习中提高。各章都有光盘数据的案例，有利于实例练习。

第三，本书是以 Microsoft Excel 带动统计计算，毋需花费时间学习 Matlab、SAS、S-Plus 等软件。通过使用 Excel 学会做统计分析，又解除了手工计算的繁琐与枯燥。

本书适合作为经济管理专业的大学生、研究生的统计学教材或参考书，也适于人文科学、社会科学、生命科学、考古学、心理学等领域的教师、工程师、技术人员自学之用。

龚光鲁

清华大学数学系

Preface

WHY I WROTE THIS BOOK

Until five years ago, virtually all applied statistics courses were taught with an emphasis on manual calculations. There were several reasons for this pedagogy. First, many instructors felt that students needed to be taught the mathematical principles of statistics, and the best way, it was thought, was to have students perform manual calculations. Second, although real statistics practitioners use a computer, costs of both hardware and software made it difficult or impossible to equip all students with these resources.

The emphasis on manual calculations has resulted in a variety of problems. Foremost among these is that for many students statistics is one of their least favorite courses. By focusing on the arithmetic of statistical techniques, instructors gave the false impression that this is the most important aspect of the subject. Students did not get the opportunity to see statistics as it really is.

This book was written to allow instructors to teach a course in applied statistics that demonstrates the true value of statistics.

APPROACH

The easy access to computers today allows us to change the pedagogy. Most students have their own computers or can use computers supplied by their colleges or universities. Moreover, nearly all computers have Microsoft Excel® automatically installed. As a result, college and university students know how to use the popular spreadsheet.

Changing the pedagogy involves far more than simply adding computer output to the illustrative examples. Fully taking advantage of these resources requires a major overhaul in the way we teach statistics. The void created by removing manual calculations can be enormous. A large portion of an instructor's class time must be devoted to showing how to calculate each statistic if the students are expected to solve problems the same way. However, the void can easily be filled by teaching aspects of applied statistics that have generally been ignored by instructors and by textbook authors. Simply expressed, we need to teach what statistics practitioners do before the calculations and what they do after the calculations, while letting the computer do the calculations themselves.

Before the calculations can be performed, the statistics practitioner must decide which technique or techniques should be applied. Without the ability to select the appropriate method, all other skills are irrelevant.

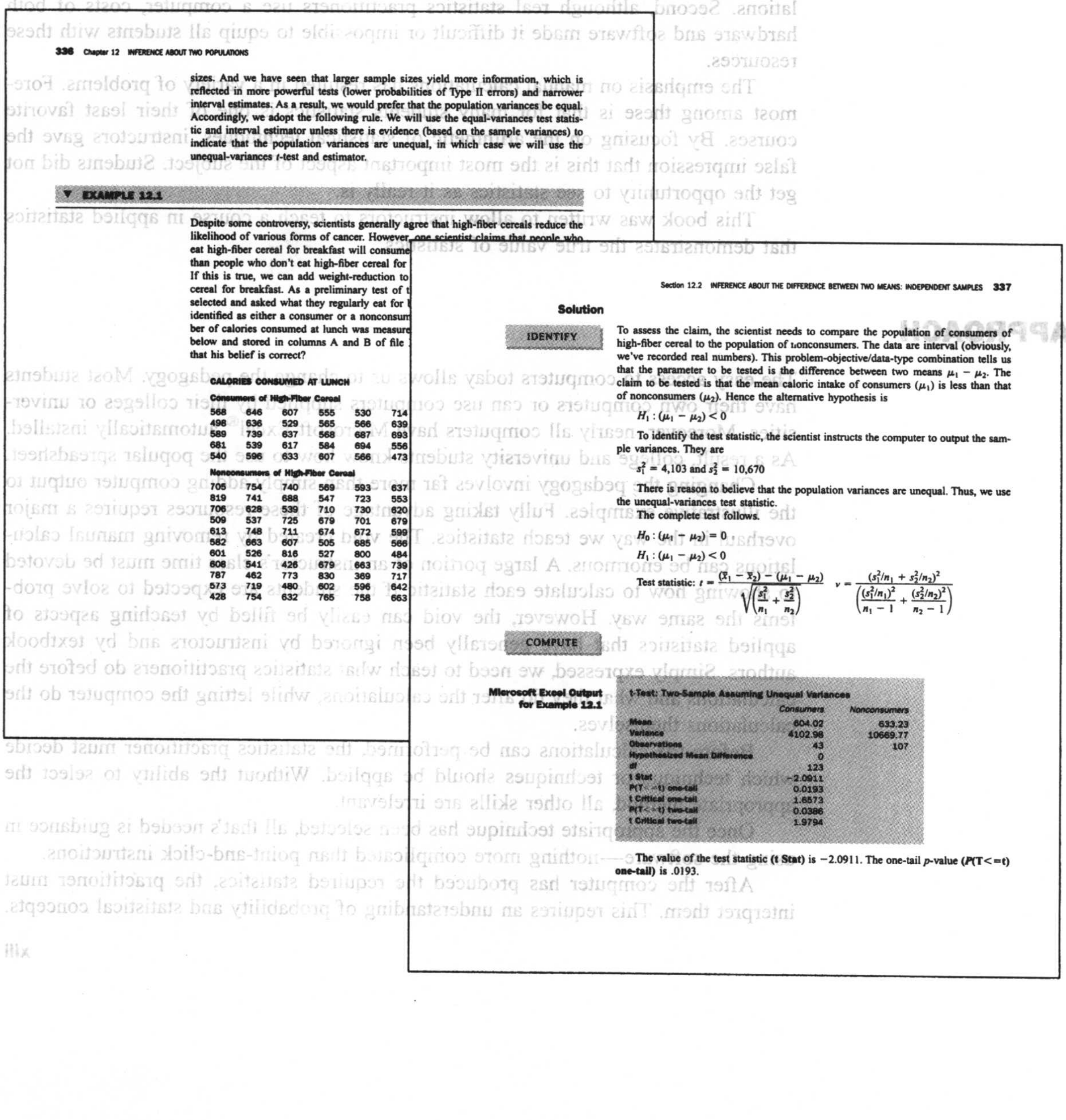
Once the appropriate technique has been selected, all that's needed is guidance in using the software—nothing more complicated than point-and-click instructions.

After the computer has produced the required statistics, the practitioner must interpret them. This requires an understanding of probability and statistical concepts.

Once again, the time needed to teach these subjects becomes available after the arithmetic is dropped from the course.

FEATURES

1. **Three-step approach.** All inferential techniques are demonstrated using a three-step approach. The first step is to identify the technique. We believe that technique-identification skills are critical but seldom taught. These skills are taught throughout the book. The second step is to compute the statistics. The third step is to interpret the results, which requires an understanding of statistical concepts.



336 Chapter 12 INFERENCE ABOUT TWO POPULATIONS

sizes. And we have seen that larger sample sizes yield more information, which is reflected in more powerful tests (lower probabilities of Type II errors) and narrower interval estimates. As a result, we would prefer that the population variances be equal. Accordingly, we adopt the following rule. We will use the equal-variances test statistic and interval estimator unless there is evidence (based on the sample variances) to indicate that the population variances are unequal, in which case we will use the unequal-variances t-test and estimator.

▼ EXAMPLE 12.1

Despite some controversy, scientists generally agree that high-fiber cereals reduce the likelihood of various forms of cancer. However, one scientist claims that people who eat high-fiber cereal for breakfast will consume than people who don't eat high-fiber cereal for If this is true, we can add weight-reduction to cereal for breakfast. As a preliminary test of t selected and asked what they regularly eat for identified as either a consumer or a nonconsum ber of calories consumed at lunch was measure below and stored in columns A and B of file that his belief is correct?

CALORIES CONSUMED AT LUNCH

Consumers of High-Fiber Cereal

568	646	607	555	530	714
498	636	529	565	566	639
589	739	637	568	687	693
681	539	617	584	694	556
540	596	633	607	566	473

Nonconsumers of High-Fiber Cereal

705	754	740	569	593	637
819	741	688	547	723	553
706	628	539	710	730	620
509	537	725	679	701	679
613	748	711	674	672	599
582	663	607	505	685	566
601	526	816	527	800	484
608	541	426	679	663	739
787	462	773	830	369	717
573	719	480	602	596	642
428	754	632	765	758	663

Section 12.2 INFERENCE ABOUT THE DIFFERENCE BETWEEN TWO MEANS: INDEPENDENT SAMPLES 337

Solution

IDENTIFY

To assess the claim, the scientist needs to compare the population of consumers of high-fiber cereal to the population of nonconsumers. The data are interval (obviously, we've recorded real numbers). This problem-objective/data-type combination tells us that the parameter to be tested is the difference between two means $\mu_1 - \mu_2$. The claim to be tested is that the mean caloric intake of consumers (μ_1) is less than that of nonconsumers (μ_2). Hence the alternative hypothesis is

$$H_1: (\mu_1 - \mu_2) < 0$$

To identify the test statistic, the scientist instructs the computer to output the sample variances. They are

$$s_1^2 = 4{,}103 \text{ and } s_2^2 = 10{,}670$$

There is reason to believe that the population variances are unequal. Thus, we use the unequal-variances test statistic.

The complete test follows.

$$H_0: (\mu_1 - \mu_2) = 0$$

$$H_1: (\mu_1 - \mu_2) < 0$$

$$\text{Test statistic: } t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} \qquad \nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\left(\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}\right)}$$

COMPUTE

Microsoft Excel Output for Example 12.1

t-Test: Two-Sample Assuming Unequal Variances

	Consumers	Nonconsumers
Mean	604.02	633.23
Variance	4102.98	10669.77
Observations	43	107
Hypothesized Mean Difference	0	
df	123	
t Stat	−2.0911	
P(T<=t) one-tail	0.0193	
t Critical one-tail	1.6573	
P(T<=t) two-tail	0.0386	
t Critical two-tail	1.9794	

The value of the test statistic (t Stat) is −2.0911. The one-tail p-value (P(T<=t) one-tail) is .0193.

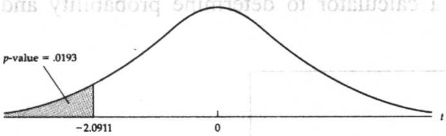


Figure 12.2
Sampling distribution of the test statistic for Example 12.1.

GENERAL COMMANDS	EXAMPLE 12.1
1 Type or import the data into two columns.	**Open file Xm12-01.**
2 Click **Tools, Data Analysis. . . ,** and **t-test: Two-Sample Assuming Unequal Variances.**	
3 Specify the **Variable 1 Range.**	**A1:A44**
4 Specify the **Variable 2 Range.**	**B1:B108**
5 Type the value of the **Hypothesized Mean Difference.**	**0**
6 Click **Labels** if applicable.	
7 Specify the value of α **(Alpha:)**, and click **OK.**	**.05**

Figure 12.2 depicts the Student t sampling distribution and the p-value of the test.

To conduct this test from means and standard deviations or to perform a what-if analysis, activate the **t-test of 2 Means (Uneq-Var)** worksheet in **Stats-Summary.xls.**

INTERPRET

A Type I error in this example occurs when we conclude that consumers of high-fiber cereal consume fewer calories at lunch than do nonconsumers when in fact there is no difference between the two groups. A Type II error occurs when we erroneously fail to conclude that consumers of high-fiber cereal eat less at lunch than nonconsumers. It is difficult to judge which error is more costly. As a result we take a neutral position and suggest that any p-value less than 5% should be interpreted as providing enough evidence to reject the null hypothesis.

The p-value of the test is .0193. As a result, we conclude that there is sufficient evidence to infer that consumers of high-fiber
than do nonconsumers. However, there are tw
cluding that high-fiber cereals constitute an effe
the data were likely self-reported, which means
ber of calories recorded that he or she consume
ally, a less subjective method of counting calori
which the experiment was performed may lead t
of the data. We will discuss this important issu

In addition to testing to determine whether a difference exists, we can estimate the difference in mean caloric intake.

The interval estimator of the difference between two means with unequal population variances is

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

COMPUTE

The Excel output does not include the interval estimate. However, you can use the **t-Estimate of 2 Means (Uneq-Var)** worksheet of **Stats-Summary.xls.** Simply plug in the values of the sample means, sample standard deviations, and sample sizes, as well as the confidence level. The result of this action appears below.

Microsoft Excel Output of the 95% confidence interval Estimate: Example 12.1

t-Estimate of the Difference between Two Means (Unequal-Variances)	
Sample 1	
Sample mean	604.02
Sample standard deviation	64.05
Sample size	43
Sample 2	
Sample mean	633.2
Sample standard deviation	103.3
Sample size	107
Confidence level	0.95
Degrees of freedom	122.60
Difference between means	−29.2
Bound	27.65
Lower confidence limit	−56.86
Upper confidence limit	−1.56

INTERPRET

We estimate that the average consumer of high-fiber cereal eats between 1.56 and 56.86 fewer calories at lunch than does the average nonconsumer of high-fiber cereal.

▼ EXAMPLE 12.2

The plant manager of a company that manufactures office equipment is attempting to determine the process that will be used to assemble a new ergonomic chair. The material, machines, and workforce have already been decided. However there are two methods under consideration. The methods differ by the order in which the separate operations are performed. To help decide which should be used, an experiment was performed. Twenty-five randomly selected workers each assembled the chair using method A, and another 25 workers each assembled the chair using method B. The assembly times in minutes were recorded and are exhibited below and stored in file Xm12-02. The plant manager would like to know whether the assembly times of the two methods differ.

2. **Fewer manual calculations.** All statistical procedures, ranging from graphical descriptive methods to multiple regression, will be conducted using a computer. We use a calculator to determine probability and to compute simple statistics.

Solution We seek

$P(X > 6{,}000)$

where X is normally distributed with $\mu = 4{,}500$ and $\sigma = 1{,}000$.

To use Table 6.1 we must standardize both X and 6,000.

$$P(X > 6{,}000) = P\left(\frac{X-\mu}{\sigma} > \frac{6{,}000 - 4{,}500}{1{,}000}\right) = P(Z > 1.5)$$

Figure 6.13 depicts this calculation. Notice that we find $P(0 < Z < 1.5) = .4332$ from Table 6.1. Because the area under the curve is 1 and half the area is above 0 (and half below), it follows that $P(Z > 0) = .5$. Thus

$$P(Z > 1.5) = P(Z > 0) - P(0 < Z < 1.5) = .5 - .4332 = .0668$$

▲

FINDING VALUES OF *z*

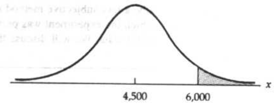
There is a family of problems that require us to determine the value of z given a probability. We use the notation z_A to represent the value of z such that the area to its right under the standard normal curve is A. That is, z_A is a value of a standard normal random variable such that

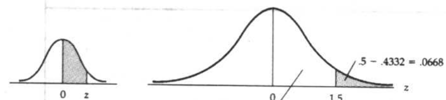
$P(Z > z_A) = A$ .025.

Figure 6.14 depicts this notation.

To find z_A for any value of A requires us to use the standard normal table backwards. As you saw in Example 6.2, to find a probability about Z we must find the value of z in the table and determine the probability associated with it. To use the table backwards we need to specify a probability and then determine the z-value associated with it. We'll demonstrate by finding z normal curve and $z_{.025}$. Because of the format o by determining the area between 0 and $z_{.025}$, wh we expressed this probability with four decima see what you need to do.) We now search thro looking for .4750. When we locate it, we see 1.96.

Thus, $z_{.025} = 1.96$, which means that $P(Z$





z	.00	.01	.02
0.0	.0000	.0040	.0080
0.1	.0398	.0438	.0478
0.2	.0793	.0832	.0871
0.3	.1179	.1217	.1255
0.4	.1554	.1591	.1628
0.5	.1915	.1950	.1985
0.6	.2257	.2291	.2324
0.7	.2580	.2611	.2642
0.8	.2881	.2910	.2939
0.9	.3159	.3186	.3212
1.0	.3413	.3438	.3461
1.1	.3643	.3665	.3686
1.2	.3849	.3869	.3888
1.3	.4032	.4049	.4066
1.4	.4192	.4207	.4222
1.5	.4332	.4345	.4357

Figure 6.13
Calculating $P(X > 6{,}000) = P(Z > 1.5)$

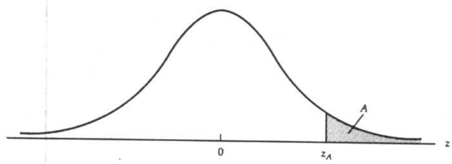
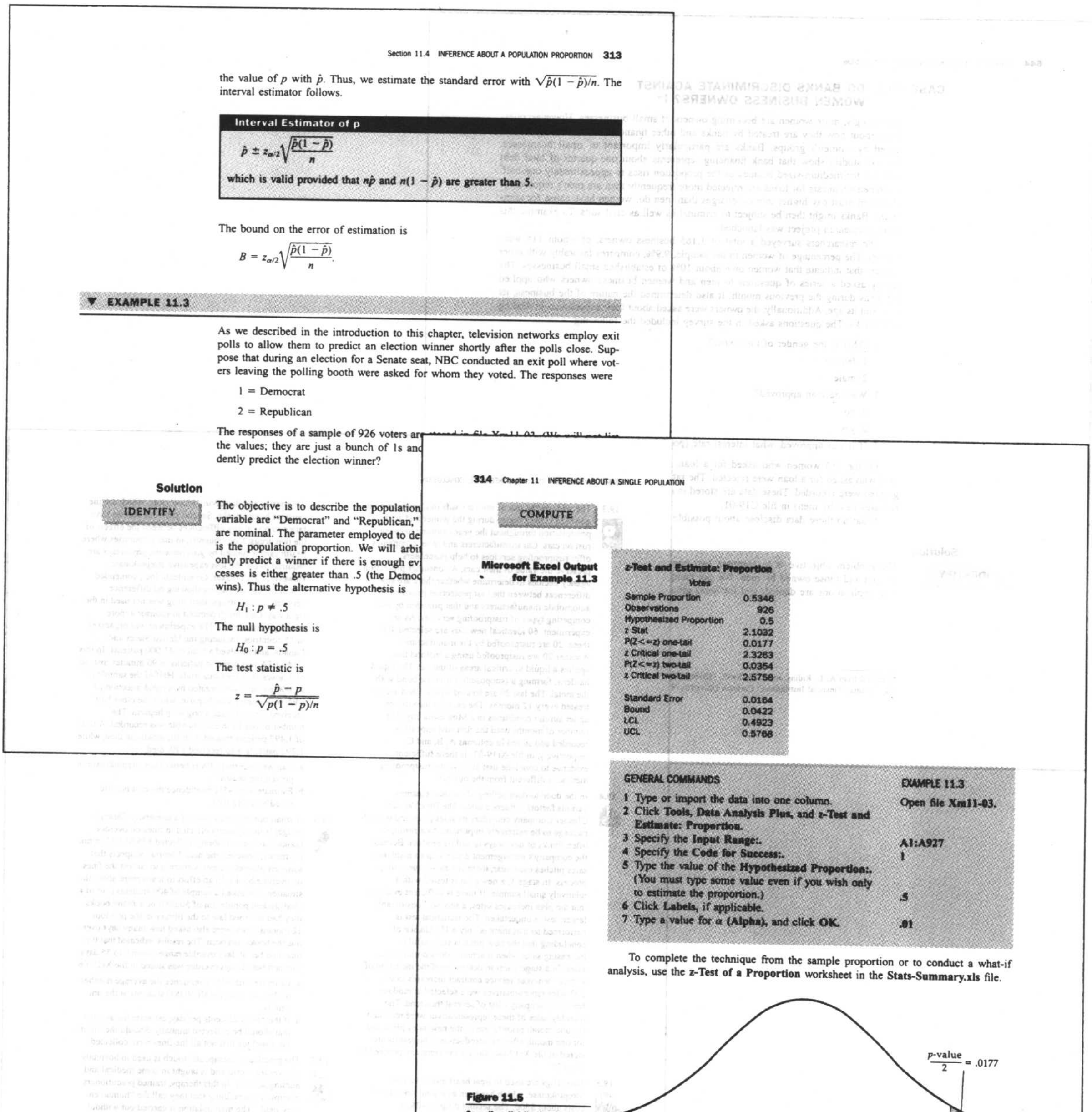


Figure 6.14
$P(Z > z_A) = A$

3. **Microsoft Excel.** We use the most popular spreadsheet program, showing the printouts and providing step-by-step instructions. We created macro add-ins to augment Excel's limited offering. The add-ins are stored on the CD, in Data Analysis Plus® 3.0, which, when installed, integrates the add-ins with Excel's statistical functions.



Section 11.4 INFERENCE ABOUT A POPULATION PROPORTION 313

the value of p with $\hat{p}$. Thus, we estimate the standard error with $\sqrt{\hat{p}(1-\hat{p})/n}$. The interval estimator follows.

Interval Estimator of p

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

which is valid provided that $n\hat{p}$ and $n(1-\hat{p})$ are greater than 5.

The bound on the error of estimation is

$$B = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

▼ EXAMPLE 11.3

As we described in the introduction to this chapter, television networks employ exit polls to allow them to predict an election winner shortly after the polls close. Suppose that during an election for a Senate seat, NBC conducted an exit poll where voters leaving the polling booth were asked for whom they voted. The responses were

1 = Democrat

2 = Republican

The responses of a sample of 926 voters ar... the values; they are just a bunch of 1s an... dently predict the election winner?

Solution

IDENTIFY

The objective is to describe the population... variable are "Democrat" and "Republican," ... are nominal. The parameter employed to de... is the population proportion. We will arbi... only predict a winner if there is enough ev... cesses is either greater than .5 (the Democ... wins). Thus the alternative hypothesis is

$$H_1: p \neq .5$$

The null hypothesis is

$$H_0: p = .5$$

The test statistic is

$$z = \frac{\hat{p}-p}{\sqrt{p(1-p)/n}}$$

314 Chapter 11 INFERENCE ABOUT A SINGLE POPULATION

COMPUTE

Microsoft Excel Output for Example 11.3

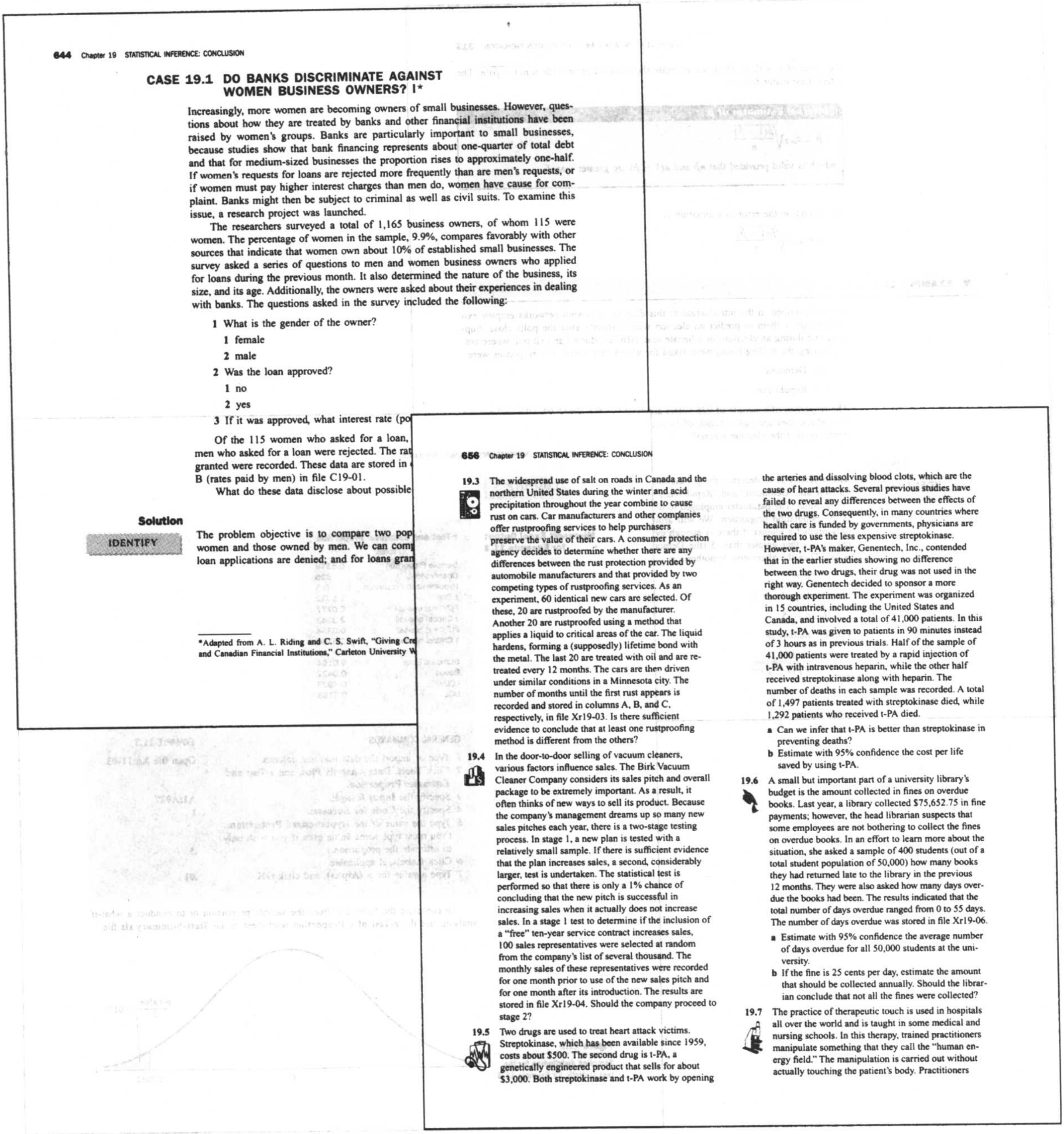
z-Test and Estimate: Proportion	
	Votes
Sample Proportion	0.5346
Observations	926
Hypothesized Proportion	0.5
z Stat	2.1032
P(Z<=z) one-tail	0.0177
z Critical one-tail	2.3263
P(Z<=z) two-tail	0.0354
z Critical two-tail	2.5758
Standard Error	0.0164
Bound	0.0422
LCL	0.4923
UCL	0.5768

GENERAL COMMANDS	EXAMPLE 11.3
1 Type or import the data into one column.	Open file Xm11-03.
2 Click **Tools, Data Analysis Plus,** and **z-Test and Estimate: Proportion.**	
3 Specify the **Input Range:.**	A1:A927
4 Specify the **Code for Success:.**	1
5 Type the value of the **Hypothesized Proportion:.** (You must type some value even if you wish only to estimate the proportion.)	.5
6 Click **Labels,** if applicable.	
7 Type a value for α **(Alpha),** and click **OK.**	.01

To complete the technique from the sample proportion or to conduct a what-if analysis, use the **z-Test of a Proportion** worksheet in the **Stats-Summary.xls** file.

Figure 11.5
Sampling distribution for the test statistic for Example 11.3

4. Emphasized applications. To motivate students and convince them of the importance of statistics, we emphasize applications. Many of the examples, exercises, and cases are adapted from real studies. Almost all feature a data set stored on the CD. There are over 600 data sets on the CD, some as large as 22,000 observations.



644 Chapter 19 STATISTICAL INFERENCE: CONCLUSION

CASE 19.1 DO BANKS DISCRIMINATE AGAINST WOMEN BUSINESS OWNERS? I*

Increasingly, more women are becoming owners of small businesses. However, questions about how they are treated by banks and other financial institutions have been raised by women's groups. Banks are particularly important to small businesses, because studies show that bank financing represents about one-quarter of total debt and that for medium-sized businesses the proportion rises to approximately one-half. If women's requests for loans are rejected more frequently than are men's requests, or if women must pay higher interest charges than men do, women have cause for complaint. Banks might then be subject to criminal as well as civil suits. To examine this issue, a research project was launched.

The researchers surveyed a total of 1,165 business owners, of whom 115 were women. The percentage of women in the sample, 9.9%, compares favorably with other sources that indicate that women own about 10% of established small businesses. The survey asked a series of questions to men and women business owners who applied for loans during the previous month. It also determined the nature of the business, its size, and its age. Additionally, the owners were asked about their experiences in dealing with banks. The questions asked in the survey included the following:

1 What is the gender of the owner?
 1 female
 2 male
2 Was the loan approved?
 1 no
 2 yes
3 If it was approved, what interest rate (po

Of the 115 women who asked for a loan, men who asked for a loan were rejected. The rat granted were recorded. These data are stored in B (rates paid by men) in file C19-01.

What do these data disclose about possible

Solution

IDENTIFY

The problem objective is to compare two pop women and those owned by men. We can comp loan applications are denied; and for loans gran

*Adapted from A. L. Riding and C. S. Swift, "Giving Cr and Canadian Financial Institutions," Carleton University W

656 Chapter 19 STATISTICAL INFERENCE: CONCLUSION

19.3 The widespread use of salt on roads in Canada and the northern United States during the winter and acid precipitation throughout the year combine to cause rust on cars. Car manufacturers and other companies offer rustproofing services to help purchasers preserve the value of their cars. A consumer protection agency decides to determine whether there are any differences between the rust protection provided by automobile manufacturers and that provided by two competing types of rustproofing services. As an experiment, 60 identical new cars are selected. Of these, 20 are rustproofed by the manufacturer. Another 20 are rustproofed using a method that applies a liquid to critical areas of the car. The liquid hardens, forming a (supposedly) lifetime bond with the metal. The last 20 are treated with oil and are re-treated every 12 months. The cars are then driven under similar conditions in a Minnesota city. The number of months until the first rust appears is recorded and stored in columns A, B, and C, respectively, in file Xr19-03. Is there sufficient evidence to conclude that at least one rustproofing method is different from the others?

19.4 In the door-to-door selling of vacuum cleaners, various factors influence sales. The Birk Vacuum Cleaner Company considers its sales pitch and overall package to be extremely important. As a result, it often thinks of new ways to sell its product. Because the company's management dreams up so many new sales pitches each year, there is a two-stage testing process. In stage 1, a new plan is tested with a relatively small sample. If there is sufficient evidence that the plan increases sales, a second, considerably larger, test is undertaken. The statistical test is performed so that there is only a 1% chance of concluding that the new pitch is successful in increasing sales when it actually does not increase sales. In a stage 1 test to determine if the inclusion of a "free" ten-year service contract increases sales, 100 sales representatives were selected at random from the company's list of several thousand. The monthly sales of these representatives were recorded for one month prior to use of the new sales pitch and for one month after its introduction. The results are stored in file Xr19-04. Should the company proceed to stage 2?

19.5 Two drugs are used to treat heart attack victims. Streptokinase, which has been available since 1959, costs about $500. The second drug is t-PA, a genetically engineered product that sells for about $3,000. Both streptokinase and t-PA work by opening the arteries and dissolving blood clots, which are the cause of heart attacks. Several previous studies have failed to reveal any differences between the effects of the two drugs. Consequently, in many countries where health care is funded by governments, physicians are required to use the less expensive streptokinase. However, t-PA's maker, Genentech, Inc., contended that in the earlier studies showing no difference between the two drugs, their drug was not used in the right way. Genentech decided to sponsor a more thorough experiment. The experiment was organized in 15 countries, including the United States and Canada, and involved a total of 41,000 patients. In this study, t-PA was given to patients in 90 minutes instead of 3 hours as in previous trials. Half of the sample of 41,000 patients were treated by a rapid injection of t-PA with intravenous heparin, while the other half received streptokinase along with heparin. The number of deaths in each sample was recorded. A total of 1,497 patients treated with streptokinase died, while 1,292 patients who received t-PA died.

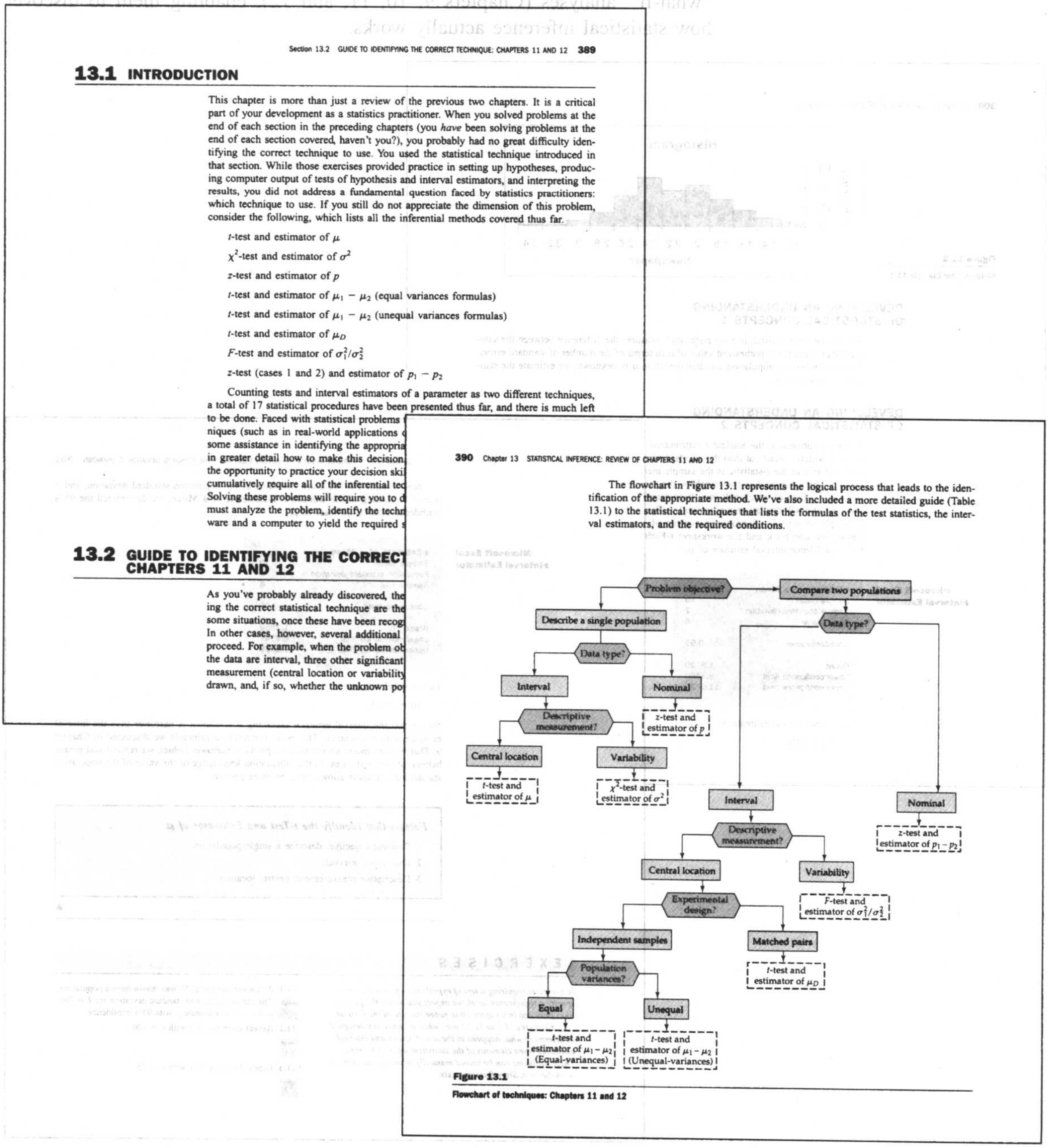
a Can we infer that t-PA is better than streptokinase in preventing deaths?
b Estimate with 95% confidence the cost per life saved by using t-PA.

19.6 A small but important part of a university library's budget is the amount collected in fines on overdue books. Last year, a library collected $75,652.75 in fine payments; however, the head librarian suspects that some employees are not bothering to collect the fines on overdue books. In an effort to learn more about the situation, she asked a sample of 400 students (out of a total student population of 50,000) how many books they had returned late to the library in the previous 12 months. They were also asked how many days overdue the books had been. The results indicated that the total number of days overdue ranged from 0 to 55 days. The number of days overdue was stored in file Xr19-06.

a Estimate with 95% confidence the average number of days overdue for all 50,000 students at the university.
b If the fine is 25 cents per day, estimate the amount that should be collected annually. Should the librarian conclude that not all the fines were collected?

19.7 The practice of therapeutic touch is used in hospitals all over the world and is taught in some medical and nursing schools. In this therapy, trained practitioners manipulate something that they call the "human energy field." The manipulation is carried out without actually touching the patient's body. Practitioners

5. **Review chapters with flowcharts.** To ensure that students can identify the correct technique, we provide two review chapters, Chapters 13 and 19, each with a flowchart that helps students choose the appropriate method. The exercises and cases in each review chapter require all the statistical techniques introduced up to that point.


Section 13.2 GUIDE TO IDENTIFYING THE CORRECT TECHNIQUE: CHAPTERS 11 AND 12 389

13.1 INTRODUCTION

This chapter is more than just a review of the previous two chapters. It is a critical part of your development as a statistics practitioner. When you solved problems at the end of each section in the preceding chapters (you *have* been solving problems at the end of each section covered, haven't you?), you probably had no great difficulty identifying the correct technique to use. You used the statistical technique introduced in that section. While those exercises provided practice in setting up hypotheses, producing computer output of tests of hypothesis and interval estimators, and interpreting the results, you did not address a fundamental question faced by statistics practitioners: which technique to use. If you still do not appreciate the dimension of this problem, consider the following, which lists all the inferential methods covered thus far.

t-test and estimator of μ

χ^2-test and estimator of σ^2

z-test and estimator of p

t-test and estimator of $\mu_1 - \mu_2$ (equal variances formulas)

t-test and estimator of $\mu_1 - \mu_2$ (unequal variances formulas)

t-test and estimator of μ_D

F-test and estimator of σ_1^2/σ_2^2

z-test (cases 1 and 2) and estimator of $p_1 - p_2$

Counting tests and interval estimators of a parameter as two different techniques, a total of 17 statistical procedures have been presented thus far, and there is much left to be done. Faced with statistical problems niques (such as in real-world applications some assistance in identifying the appropria in greater detail how to make this decision the opportunity to practice your decision ski cumulatively require all of the inferential te Solving these problems will require you to d must analyze the problem, identify the techn ware and a computer to yield the required s

13.2 GUIDE TO IDENTIFYING THE CORRECT CHAPTERS 11 AND 12

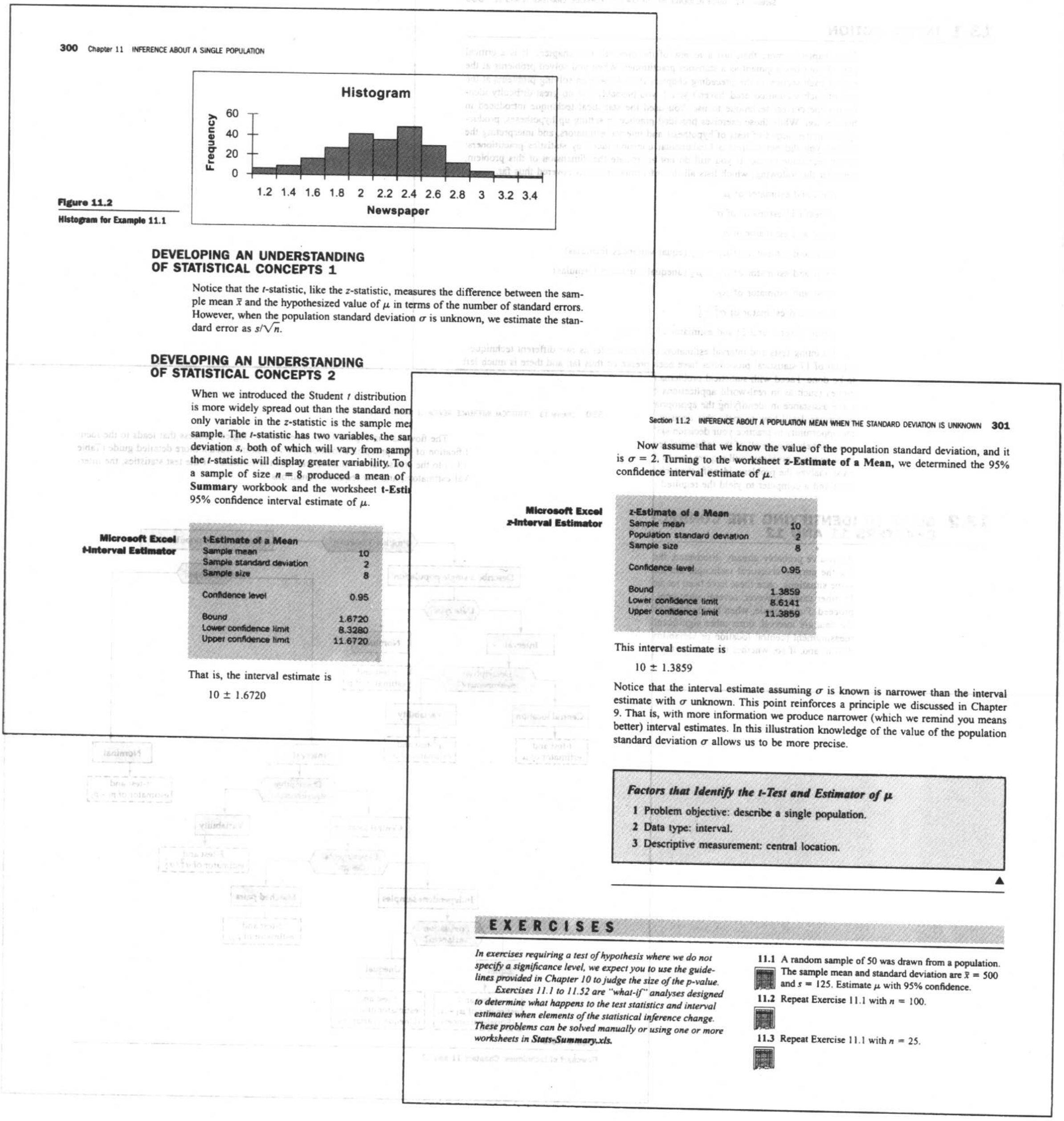
As you've probably already discovered, the ing the correct statistical technique are th some situations, once these have been recog In other cases, however, several additional proceed. For example, when the problem ob the data are interval, three other significant measurement (central location or variabilit drawn, and, if so, whether the unknown po

390 Chapter 13 STATISTICAL INFERENCE: REVIEW OF CHAPTERS 11 AND 12

The flowchart in Figure 13.1 represents the logical process that leads to the identification of the appropriate method. We've also included a more detailed guide (Table 13.1) to the statistical techniques that lists the formulas of the test statistics, the interval estimators, and the required conditions.

Figure 13.1
Flowchart of techniques: Chapters 11 and 12

6. **Developed concepts.** In the absence of mathematical derivations, it is necessary to emphasize the mathematical principles underlying all inferential methods. We do this in a variety of ways, including simulations. We describe how to simulate sampling distributions (Chapter 8) and interval estimators (Chapter 9). A spreadsheet—also stored on the CD—allows students to conduct "what-if" analyses (Chapters 9, 10, 11, and 12), enabling them to discover how statistical inference actually works.



300 Chapter 11 INFERENCE ABOUT A SINGLE POPULATION

Figure 11.2
Histogram for Example 11.1

DEVELOPING AN UNDERSTANDING OF STATISTICAL CONCEPTS 1

Notice that the t-statistic, like the z-statistic, measures the difference between the sample mean $\bar{x}$ and the hypothesized value of μ in terms of the number of standard errors. However, when the population standard deviation σ is unknown, we estimate the standard error as $s/\sqrt{n}$.

DEVELOPING AN UNDERSTANDING OF STATISTICAL CONCEPTS 2

When we introduced the Student t distribution ... is more widely spread out than the standard nor... only variable in the z-statistic is the sample me... sample. The t-statistic has two variables, the sa... deviation s, both of which will vary from samp... the t-statistic will display greater variability. To ... a sample of size $n = 8$ produced a mean of ... **Summary** workbook and the worksheet **t-Esti**... 95% confidence interval estimate of μ.

Microsoft Excel t-Interval Estimator

t-Estimate of a Mean	
Sample mean	10
Sample standard deviation	2
Sample size	8
Confidence level	0.95
Bound	1.6720
Lower confidence limit	8.3280
Upper confidence limit	11.6720

That is, the interval estimate is

10 ± 1.6720

Section 11.2 INFERENCE ABOUT A POPULATION MEAN WHEN THE STANDARD DEVIATION IS UNKNOWN **301**

Now assume that we know the value of the population standard deviation, and it is $\sigma = 2$. Turning to the worksheet **z-Estimate of a Mean,** we determined the 95% confidence interval estimate of μ.

Microsoft Excel z-Interval Estimator

z-Estimate of a Mean	
Sample mean	10
Population standard deviation	2
Sample size	8
Confidence level	0.95
Bound	1.3859
Lower confidence limit	8.6141
Upper confidence limit	11.3859

This interval estimate is

10 ± 1.3859

Notice that the interval estimate assuming σ is known is narrower than the interval estimate with σ unknown. This point reinforces a principle we discussed in Chapter 9. That is, with more information we produce narrower (which we remind you means better) interval estimates. In this illustration knowledge of the value of the population standard deviation σ allows us to be more precise.

Factors that Identify the t-Test and Estimator of μ

1. Problem objective: describe a single population.
2. Data type: interval.
3. Descriptive measurement: central location.

EXERCISES

In exercises requiring a test of hypothesis where we do not specify a significance level, we expect you to use the guidelines provided in Chapter 10 to judge the size of the p-value.

Exercises 11.1 to 11.52 are "what-if" analyses designed to determine what happens to the test statistics and interval estimates when elements of the statistical inference change. These problems can be solved manually or using one or more worksheets in ***Stats-Summary.xls.***

11.1 A random sample of 50 was drawn from a population. The sample mean and standard deviation are $\bar{x} = 500$ and $s = 125$. Estimate μ with 95% confidence.

11.2 Repeat Exercise 11.1 with $n = 100$.

11.3 Repeat Exercise 11.1 with $n = 25$.

MICROSOFT EXCEL

Microsoft Excel is a spreadsheet package that performs *some* statistical analyses. Nevertheless, we chose Excel instead of a full-feature statistical package like SAS or SPSS, or a student-friendly package like Minitab, because of the problem of computer accessibility.

The high cost of SAS, SPSS, and other software requires universities to provide the statistical program on their computer systems. However, almost all universities' computer facilities are overcrowded at the best of times and heavily congested when assignments for required courses are due. This circumstance makes the frequent use of a campus computer difficult and likely to be frustrating for students. Consequently, a virtual prerequisite to using computers in a statistics course is for students to own their own computers and software.

Fortunately, an increasing proportion of students have their own computers. Moreover, many students own a version of Excel and know how to use it. Thus, Microsoft Excel was chosen as the only viable contender. Once chosen, it became one of our tasks to ensure that all the statistical techniques presented in this book could be performed with Excel. The creation of Data Analysis Plus® (a series of add-ins, now in version 3.0) fulfills this goal.

TEACHING AND LEARNING AIDS

Instructor's Solutions Manual
This manual supplies complete solutions for every exercise in the book (ISBN: 0-534-38204-5).
Student Solutions Manual
This manual, for sale to students, provides the solutions for the textbook's even-numbered exercises (ISBN: 0-534-98202-9).

ACKNOWLEDGMENTS

I would like to express my appreciation to Carolyn Crockett, the book's editor, whose advice and encouragement improved both the quality and speed of the project. Margaret Pinette and Tessa Avila did the production work, and Seema Atwal assisted with the supplementary material. I also want to thank Jeffery Keller and Brendan Paul, who produced the latest version of Data Analysis Plus®.

The following reviewers contributed helpful feedback on the manuscript: Michael Baron, The University of Texas at Dallas; John Dutton, North Carolina State University; James Guffey, Truman State University; Susan Jenkins, Southeastern Louisiana University; Robert Lee, James Madison University; and Carolyn Meitler, Concordia University Wisconsin.

Gerald Keller

KEY TO APPLICATION ICONS

 Calculations

 Definitions

 Everyday life

 Education

 Business/Economics

 Medicine/Health

 Agriculture

 Environment

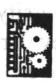
 Internet

 Television

 Public opinion

 Politics

 Sports/Recreation

 Engineering/Technology

 Science

 Gambling