Decisions! Decisions! Decisions! Do you attend Harvard University or KITTENS University? Do you marry this person or not? Does your pizza company continue the television advertising which features the “Pizza for People Who Like to Canoodle” slogan?

Success in life is 90% making the right decisions in the first place. And only 10% carrying out those decisions.

People with good decision-making skills are rare. They are also the most valuable persons in any business, any army or any orchestra. These CEOs, generals and conductors all have the same job: they take massive amounts of data and boil them down to yes-or-no decisions.

- Shall we sell all the stock we own? (It’s September 1929.)
- Shall we launch the invasion today? (It’s June 6, 1944.)
- Shall we send our orchestra on a worldwide tour this month? (It’s early December 1941.)

And where there are numbers involved, statistics is an important aid in making good decisions. At its best, statistics is a way of melting down a heap of numerical data into a simple yes or no. It’s a way of getting rid of numbers!

If you really hate to see big piles of numbers, you and statistics were made for each other.
A Note to Students

One morning in the life of Fred. A Saturday just after his sixth birthday. In his everyday life Fred will run into the need for every kind of statistics. Each time we do a little statistics, we see how it helps him get through his morning.

HOW MUCH STATISTICS IS COVERED IN THIS BOOK?

We start at the beginning with simple descriptive statistics (averages, standard deviation, etc.) and then do some probability, including conditional probability with Bayes’ Theorem.

Next comes inferential statistics—the heart of statistics—in which we study a zillion* different procedures. We describe each in detail and tell you when and where each test is appropriate. You get plenty of worked-out examples for each test.

All the popular tests such as the Normal Distribution and the Chi-Squared Test are included. Many advanced tests such as the Kolmogorov-Smirnov Test and the Two-Factor ANOVA for multiple observations per cell are covered. When the Chi-Squared Test won’t work because the sample sizes are too small, we turn to Fisher’s Exact Test. Most beginning statistics books don’t include that test.

We have one test that no other statistics book mentions—at least not until future authors copy it out of this book. It deals with a small sample from a binomial distribution. Suppose, for example, a new species of fish is discovered in the ocean and of the first ten caught, three had red fins. What is the number of red-finned fish you might expect if you caught 10,000? [Answer: 95% of the time, you would expect between 1093 and 6096.] This question would stump most statistics teachers (who don’t have a copy of this book).

After the descriptive and inferential statistics, we spend the last hour or so of Fred’s morning working with regression equations including nonlinear curve fitting and logistic regression.

This book has much more material than is normally covered in a beginning university statistics course.

* 46 by actual count
HOW THEORETICAL IS THIS BOOK?

Life is practical. This is a book that will teach you how to do statistics—lots of it. Even if you are going to get a Ph.D. in statistics and are dying to go through tons of theory and proofs, your first logical step should be to learn how to do the various tests. Then, in a later course, the proofs would be appropriate. In beginning algebra, for example, you were first told that a negative number times a negative number gives a positive answer. Later, you may have seen the proof.

In this book you learn how to perform the Kruskal-Wallis Test for three or more independent samples, but we’re not going to fill up the pages with a proof.

There are two exceptions. The first is a little three-line proof of Bayes’ Theorem, which is so cute that I couldn’t resist including it. And the second is the underpinnings of the SMALL SAMPLE FROM A BINOMIAL DISTRIBUTION TEST that I mentioned on the previous page. Since no other book has this test, I placed this material in its own separate little chapter (Chapter 5½) and laid out the reasoning to show why this test works. This little chapter is the only place in the book in which there is any calculus. And even here, the calculus is very basic. It deals with the area under a curve described by a polynomial. If you go directly from Chapter 5 to Chapter 6 and bypass Chapter 5½, you will be protected from all calculus.

In doing their proofs, some books go nuts with subscripts and primes and “hats” and Greek letters. They wind up with expressions like $\hat{y}_{i,j} + \epsilon$, which certainly don’t help anyone’s digestion. Those things are kept to a minimum in Life of Fred: Statistics. [$\hat{y}$ is read “y-hat.”]

WHAT BACKGROUND DO I NEED?

It would be nice to have a little algebra so that $x^2$ and absolute values and square roots don’t mystify you. But that’s about it. I can’t think of anywhere in the book where you’ll need to solve any equations or do any algebra word problems.**

We’ll use the greater than sign (>) and plus-or-minus (±).

🌟 None of those old word problems like: JACKIE IS CHASING D ALE DOWN THE HALL WITH AN AX. JACKIE IS TRAVELING 7 FT/SEC AND D ALE IS RUNNING AT 5 FT/SEC. THEY ARE 8 FEET APART. HOW SOON SHOULD D ALE START APOLOGIZING?
See if these all make sense to you:

- $7^2 = 49$
- $|-3| = 3$
- $64 > 29$
- $7 \pm 2$ means $5$ or $9$.
- Using your calculator $\sqrt{3}$ gives $1.7320508$.

If so, you are ready.

**DO I NEED A COMPUTER?**

No.

**DO I NEED A GRAPHING CALCULATOR?**

No. All you need is a handheld calculator that has keys like $\sin$, $\cos$, and $\log$. Those calculators don’t cost that much. Certainly under $20$. (I have seen them under $10$.) In a couple of years they will probably be included free in cereal boxes.

**ANY SPECIAL SUGGESTIONS BEFORE I START CHAPTER ONE?**

Yes. I have a couple of ideas.

First, in each chapter there are *Your Turn to Play* sections. These have representative problems along with completely worked-out solutions. Please solve these problems before you glance at the solutions. Just reading the problems and eyeballing the solutions is a real temptation for some readers, but unless you’re smarter than Einstein, you won’t learn much doing that.

At the end of each chapter are six sets of exercises which I call Cities. The first two Cities have all the answers supplied. The second pair of Cities have answers to all the odd-numbered problems. If you want to learn statistics, please do all these problems for which I have given the answers.

Second, I need to know if you are in a real hurry.
If that’s the case, then don’t start by turning to the first page of Chapter One, or to the Table of Contents or to the Index.

Instead, turn to the Emergency Statistics Guide which begins on p. 326. The Emergency Statistics Guide will tell you:

① what test to use,

② where to find an explanation of the test as it occurred in Fred’s life,

③ where it’s listed in the Field Guide and

④ what table to use.

The Emergency Statistics Guide will move you from baffled to brilliant in twelve seconds flat.
Contents

Chapter 1  Descriptive Statistics ................................. 21
  frequency distributions
  scatter diagrams
  averages—mean, median and mode
  linear regression
  populations vs. samples
  histograms
  range
  percentiles, deciles, quintiles, quartiles
  variance
  sigma notation
  standard deviation for populations and for samples
  distributions—skewed, platykurtic, leptokurtic, bimodal

Chapter 2  Probability ........................................... 56
  outcomes
  sample space
  events—Independent, complements, mutually exclusive
  Venn diagrams

Chapter 3  Conditional Probability ............................... 72
  \(P(A | B)\) notation
  definition of conditional probability
  Bayes’ Theorem and its proof
  generalized Bayes’ Theorem

Chapter 3½ Looking Forward to the Next Four Chapters .......... 91
  the Future—zero samples
  the Past—one sample
  the Present—two samples
  the Present—three or more samples

Chapter 4  The Future—Zero Samples ............................ 94
  Poisson distributions
  factorial
  continuous vs. discrete variables
  exponential distributions—three forms
  permutations and combinations
  Bernoulli variables
  binomial distributions
  hypergeometric distributions
  multinomial distributions
Chapter 4½  The Art of the Sample .......................... 124
null hypothesis—$H_0$
the problem of induction—Hume’s problem
the problem of small samples
type I and type II errors
levels of significance
The Ten Rules of Fair Play
data mining, cherry picking, data snooping
pilot samples
alternative hypotheses
one-tail vs. two-tail propositions
dealing with sensitive questions in a survey
dealing with bad luck in surveys
simple random surveys
systematic samples
cluster sampling
stratified samples
outliers
statistical significance vs. actual significance
13 alternatives to saying “$H_0$ is tenable.”

Chapter 5  The Past—One Sample .......................... 145
why no one knows what time it is
Normal Distributions—large samples, but a small part of the population
z-scores
determining sample size
confidence intervals
Central Limit Theorem
point estimates
Wald confidence intervals vs. Agresti-Coull confidence intervals
finite population correction factors
Normal Distributions—large samples that are a large part of the population
Student’s $t$-Distribution
Lilliefors test for normality
standardizing data
cumulative normal frequency
Wilcoxon Signed Ranks test—the Median test
uniform distributions
symmetric distributions
Sign test
power of a test
data—nominal, ordinal, interval, ratio
parametric vs. nonparametric statistics
Sign test for nominal data
Kolmogorov-Smirnov goodness-of-fit test
  for uniform distributions
  for normal distributions
Chi-Squared test
  for goodness-of-fit test
  the Lie Detector test
  is-the-sample-too-variable test
sequences—random, cyclical, trends
Runs test

Chapter 5½  Secrets of the Binomial Proportion ................. 204
  starting with a small sample of a Bernoulli variable
  we determine the confidence interval for $\pi$, the proportion
  of “good” items in the underlying population
  a small history of the problem
  Monte Carlo method
  the journal article (from The Journal of Freedometrika),
    which describes a new approach to the problem

Chapter 6  The Present—Two Samples ......................... 214
  paired samples
  Two Paired Samples ($\mu_1 - \mu_2$) test
  Wilcoxon Signed Ranks test for two paired samples
  Signs test for two paired samples
  Signs test for paired samples of Hot & Cold
  Two Proportions with 2 samples in 2 categories
  independent samples
  Two Large Independent Samples test
  Two Independent samples when $\sigma_1$ and $\sigma_2$ are known
  F-Distribution test
  Two Small Independent Samples test where the populations
    are normal and the standard deviations are roughly equal
  Two Small Independent Samples test where the populations
    are roughly normal but the standard deviations are quite
    different from each other (a. k. a. the Smith-Satterthwaite
    test)
  Mann-Whitney test
  Chi-Squared test for 2 samples in many categories
  contingency tables
  one sample with two variables
  Chi-Squared test with Yates Correction for 2 samples in 2
    categories
  Fisher’s Exact test for 2 samples in 2 categories
Chapter 7 The Present—Many Samples .......................... 266
One-Way ANOVA test for independent samples
weighted averages
Post-test for One-Way ANOVA for independent samples
One-Way ANOVA for matched samples (blocked samples)
Post-test for One-Way ANOVA for matched samples
Two-Factor ANOVA with one observation per cell
Post-test for Two-Factor ANOVA
ANOVA tables
Two-Factor ANOVA with several observations per cell
Kruskal-Wallis test
Post-test for Kruskal-Wallis
Chi-Squared test for nominal data, three or more samples
correlation vs. causation

Chapter 7½  Emergency Statistics Guide  ......................... 326

Chapter 8 Finding Regression Equations ....................... 342
linear regression
prediction intervals
Pearson Product Moment Correlation Coefficient (r)
coefficient of determination
multiple regression
normal equations
coefficient of multiple determination (R²)
adjusted coefficient of multiple determination
design variables, dummy variables
saturated models
multicollinearity method
step down method
nonlinear regression
  logarithmic curves
  reciprocal curves
  power curves
  exponential curves
  parabolic curves
two independent variables with possible interaction
logistic regression

The Field Guide ...................................................... 388
Future—The population is known and you want to know what the sample will look like. You start with zero samples.
Hypergeometric Distribution
Extended Hypergeometric Distribution
Binomial Distribution
Multinomial Distribution
Poisson Distribution

/0
Exponential Distribution
Normal Distribution

Past—The sample is known and you want to know what the population was that gave this sample. You start with one sample.
Normal Distribution—n > 30 and the sample is small compared with the population.
Normal Distribution—n > 30 and the sample is large compared with the population.
Student’s t-Distribution
Binomial Distribution (large sample, n > 30)
Binomial Distribution (small sample, n ≤ 30)
Kolmogorov-Smirnov goodness-of-fit test
Lilliefors test
Wilcoxon Signed Ranks test
Sign test—Does the population have that median?
Sign test for Nominal Data
Chi-Squared test (goodness of fit)
Chi-Squared test (Lie Detector)
Chi-Squared test (Is the population too variable?)
Runs test

Present—You start with two samples and want to know how do they compare with each other.
Two Paired Samples (μ₁ – μ₂)
Wilcoxon Signed Ranks test
Sign test for two paired samples
Sign test for two paired samples of nominal data.
Two Proportions in two categories.
Two Large Independent Samples, n ≥ 30
Two Independent Samples (σ₁ and σ₂ known)
F-Distribution test
Two Small Independent Samples, roughly equal standard deviations
Two Small Independent Samples (Smith-Satterthwaite) with very different standard deviations.
Mann-Whitney test (a.k.a. Wilcoxon Rank-Sum test)
Chi-Squared test (χ²), two samples of nominal data in multiple categories.
One Sample with Two Variables
Chi-Squared test (χ²) — Yates Correction
Fisher’s Exact test

Present—You start with three or more samples and want to know how do they compare with each other.
One-Way ANOVA (independent samples)
Post-test for One-Way ANOVA (independent samples)
One-Way ANOVA (matched samples)
Post-test for One-Way ANOVA (matched samples)
Two-Factor ANOVA (one observation per cell)
Post-test for Two-Factor ANOVA (one observation per cell)
Two-Factor ANOVA (multiple observations per cell)
Kruskal-Wallis test
Post-test for Kruskal-Wallis
Chi-Squared ($\chi^2$), three samples of nominal data

## Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Binomial Coefficients</td>
</tr>
<tr>
<td>B</td>
<td>Kolmogorov-Smirnov (one sample)</td>
</tr>
<tr>
<td>C</td>
<td>Standard Normal Curve (area from 0 to $z$)</td>
</tr>
<tr>
<td>D</td>
<td>Standard Normal Curve (area from $-\infty$ to $z$)</td>
</tr>
<tr>
<td>E</td>
<td>Standard Normal Curve (area from $-z$ to $z$)</td>
</tr>
<tr>
<td>F</td>
<td>Student’s $t$-Distribution</td>
</tr>
<tr>
<td>G</td>
<td>Lilliefors</td>
</tr>
<tr>
<td>H</td>
<td>Wilcoxon Signed Ranks</td>
</tr>
<tr>
<td>I</td>
<td>Sign test</td>
</tr>
<tr>
<td>J</td>
<td>Chi-Squared ($\chi^2$)</td>
</tr>
<tr>
<td>K</td>
<td>Runs test</td>
</tr>
<tr>
<td>L</td>
<td>Mann-Whitney (Wilcoxon Rank-Sum)</td>
</tr>
<tr>
<td>M</td>
<td>Fisher’s Exact test</td>
</tr>
<tr>
<td>N</td>
<td>$F$-Distribution</td>
</tr>
<tr>
<td>O</td>
<td>Kruskal-Wallis test</td>
</tr>
<tr>
<td>P</td>
<td>Binomial Proportion Intervals</td>
</tr>
</tbody>
</table>

## Index

* 539
ink! Fred’s eyes popped open. He had just heard one of the sweetest sounds. He looked at his watch. 4:13 A.M. With his mouth open, he listened in the dark. Tink! Yes, he thought to himself, it’s happened. Tink!

Drops of water were falling from the ceiling. Fred threw off his bedcovers and emerged from under his desk. He looked at the pot on his desktop and saw three drops of water. Tink! Make that four drops.

His watch clicked over to 4:14 A.M. and he smiled as six more drops fell into the pot. It’s a little early to telephone Alexander, Fred thought, but it won’t hurt if I email him. Fred rolled up his three-foot sleeping bag and put it in the closet. He turned on the computer, changed out of his pajamas, turned off his nightlight, and looked out the window. From the window in his office/home he could look out over the university campus. For the first time in months, the sky was inky black and filled with stars. It was a welcome change from what he called the “dodo bird” sky of Kansas in winter.

From September through May, the cloud cover always reminded Fred of the soft, gray feathers of that extinct bird.

He opened the window and felt a warm breeze. So much to be grateful for. I teach at a wonderful university. I have my health. I have wonderful friends like Alexander and Betty. Fred uttered the prayer that God most likes to hear (“Thank you”) and then turned to his computer that was in the final stages of booting up. He put three phone books on a chair and hopped on top of them. When you’re only six years old and 36 inches tall, you need to make those kinds of adjustments in order to sit at a big-people’s desk.

On a clipboard he wrote out a little frequency distribution showing the data he had collected so far:

<table>
<thead>
<tr>
<th>time</th>
<th>no. of drops</th>
</tr>
</thead>
<tbody>
<tr>
<td>4:13</td>
<td>4</td>
</tr>
<tr>
<td>4:14</td>
<td>6</td>
</tr>
</tbody>
</table>

But that looked much too “numbery” for Fred’s taste. He liked to keep things simple. Instead of
Chapter One  Descriptive Statistics

4:13 A.M., Fred wrote “1” to stand for the first minute of spring, and “2” for the second minute. His frequency distribution looked much nicer now:

<table>
<thead>
<tr>
<th>time</th>
<th>no. of drops</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>

He stared at the computer screen. Three operating systems had been loaded, the anti-virus program and the anti-spam programs were activated, and the screen colors were being adjusted to match the university colors, and now the Internet service provider was being dialed.

Fred had a very new machine (it was a gift from his students), but the university had very old phone lines. “ISP IS NOT RESPONDING” appeared on his screen. “ERROR 397 THE NUMBER IS BEING REDIALED.”

Fred went back to looking at the pot. It was 4:19 A.M. and during that minute Fred counted 16 drops coming from the ceiling into his pot. His screen flashed, “LOCAL NUMBER IS UNAVAILABLE. THE NEVADA NUMBER IS BEING DIALED.” Fred went back to counting. Twenty drops came in the next minute. “THE NEVADA NUMBER IS BUSY. URUGUAY IS BEING DIALED.”

Fred went back to his clipboard and expanded his frequency distribution:

<table>
<thead>
<tr>
<th>time</th>
<th>no. of drops</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>7</td>
<td>20</td>
</tr>
</tbody>
</table>

To pass the time waiting for his computer, he drew a little graph.

A bunch of dots on a graph (where paired observations are plotted) is called a scatter diagram.)
Quick! No time to wade through a table of contents or an index. Do you use the Kolmogorov-Smirnov one-sample test or do you perform a Chi-Squared test? Should you resort to the Wilcoxon Signed Ranks test? Or two-factor ANOVA?

Just answer these questions and follow the arrows. You’ll learn exactly which statistics procedure you’ll need.

Do you really want to mess with the numbers and formulas yourself?

Yes

You have a population and you have sample(s) drawn from that population.

- You might be just starting with a population and are wondering what samples from that population will look like.

- You have a single sample and you want to describe the population from which it might have come.

- You have two samples and you want to compare them.

- You have three or more samples and you want to see how they fit together.

No

Hiring a statistician is an alternative. Please have your checkbook ready.

Go to next page.

Turn to page 329.

Turn to page 333.

Turn to page 339.
From the previous page.
You have a population and you want to know what samples from this population will look like.
We call this the Future.
You are starting with zero samples.

Are the data nominal/ordinal or are they interval/ratio? (The types of data—nominal, ordinal, interval and ratio—are all explained on p. 176)

nominal or ordinal data

Are you dealing with “hits” scattered randomly throughout an interval (of time, distance, etc.) or is it more like drawing colored marbles out of a jar?

hits in an interval

Do you want to find the probability of getting  g  hits in the next five minutes (or five miles or five gallons or five acres) or do you want to find the probability of getting the next hit in the next five minutes?

interval or ratio

Normal Distribution
Explanation ··· p. 116
Field Guide . . . p. 399
Table . . . . . . p. 470

Poisson Distribution
Explanation ······· p. 95
Field Guide ······· p. 399

Exponential Distribution
Explanation ······· p. 100
Field Guide ······· p. 400
Is the population that you are sampling finite or is it so large that it is effectively infinite?

- finite

infinite or pretty close to infinite

Are there just two categories, such as red and green marbles, or are there more than two kinds of marbles?

- 2

3 or more

Are you sampling with replacement? That’s the same as asking if the probabilities stay the same as you select item after item from the population.

- YES

NO

Are there just two categories, such as red and green marbles, or are there more than two kinds of marbles?

- 2

3 or more

From the previous page.

You are drawing colored marbles out of a jar.

---

- **Binomial Distribution**
  - Explanation . . . p. 105
  - Field Guide . . . p. 397
  - Table . . . . . . p. 461

- **Multinomial Distribution**
  - Explanation . . . p. 108
  - Field Guide . . . p. 398

- **Hypergeometric Distribution**
  - Explanation . . . p. 106
  - Field Guide . . . p. 395
  - Table . . . . . . p. 461

- **Hypergeometric Distribution (extended)**
  - Explanation . . . p. 110
  - Field Guide . . . p. 396
  - Table . . . . . . p. 461
Field guides can be fun. When you’re out tromping in the woods and spot this large, flightless bird. You want to know what it is. Your hiking companion remarks, “It ain’t a chicken, but it looks like good eating.”

You pull out your Acme Field Guide to Birdies and read aloud, “RAPHUS CUCULLATUS: length is about three feet. Soft, gray feathers. Forest-dwelling. Female lays one egg in a bed of grass. Used to live in Mauritius.”

“That’s fine,” your friend says as he takes aim with a rock. “That thing can’t even fly. Did you bring your cookbook?”

“Stop!” you exclaim. “We really don’t want to kill that dodo. My field guide says that they became extinct in 1693. We could make a zillion dollars if we capture this guy alive.”

See how handy a field guide can be?

And if you had a field guide for geography, you could locate Mauritius. It’s east of Madagascar. And Madagascar is a large island in the Indian Ocean. And the Indian Ocean is east of the southern part of Africa.

All of which forms the perfect segue (SEG-way, a smooth transition) to our Field Guide to Statistics. Our “birds” are divided into three different types. These three types of statistical maneuvers are called the Future, the Past, and the Present. Each of them will answer a particular type of question.

The Future begins with a known population. With these procedures we will be able to predict what a sample from this population will look like.

The Past begins with a given sample. We look at the sample and try to describe what was the population that generated that sample.

The Present begins with two or more samples and learns how these samples fit in the general scheme of things.
# Fisher’s Exact Test

## Type of Test
- **“Future”—** Population known. What *will* the sample be like?
- **“Past”—** Sample known. What *was* the population that gave this sample?
- **“Present”—** Multiple samples. How *do* they fit?

## Type of data
- **Nominal**—3 wins, 2 loses
- **Ordinal**—freshman, sophomore, junior, senior
- **Interval**—temperature
- **Ratio**—$6, \$7, \$23

## Special Features
Fisher is used when the Chi-Squared won’t work because the samples are too small—even after you have combined samples and categories down to a $2 \times 2$ contingency table.

## Variables You Know
In attempting to use the $\chi^2$ test, you have combined samples and categories and still have too few items of data in a $2 \times 2$ contingency table:

<table>
<thead>
<tr>
<th></th>
<th>w</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td></td>
<td>z</td>
</tr>
</tbody>
</table>

w, x, y and z are numbers

## What You Can Find Out
What is the probability that the two populations are alike?

## The Procedure

1. Interchange the rows and/or interchange the columns until the number in the upper left-hand box is the smallest of the four numbers.
2. If the number in the upper right-hand box is greater than the number in the lower left-hand box, interchange the two numbers.
3. Your contingency table will now look like:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

With these values of A, B, C and D, consult Table M (on p. 499) for the exact probability that the populations are the same.

## Example
In competition, St anthony’s Veal Pizza has won 6 Grand awards and lost 3 times. His Linguini-and-Antelope Pizza has won 2 Grand awards and lost 8 times. At the 5% significance level are these two pizzas different in their likelihood of winning awards?

<table>
<thead>
<tr>
<th></th>
<th>6</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Interchanging rows:

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Interchanging upper right/lower left:

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

So A = 2, B = 6, C = 8, and D = 3. Table M states there is a 0.0549 probability that the samples were drawn from the same population. That exceeds 5%. At the 95% confidence level we could not say that the two pizzas differed in their likelihood of winning awards.
adjusted coefficient of multiple determination ........ 354
Agresti-Coull confidence interval ........ 153
ANOVA table .................. 289
antilog ................. 369
average
mean .................. 31
median .................. 39
mode .................. 37
bad luck .................. 136
Bayes’ Theorem .................. 82
generalized .................. 86
proof .................. 83
bell-shaped curve ............. 49
Bernoulli variable .......... 104, 204, 373
bimodal .................. 49
binomial distribution ........ 105, 397
for proportions ............ 152
small sample ............ 204
Central Limit Theorem ........ 150
Chi-Squared test
combining categories ........ 189
goodness-of-fit test ........ 187, 416
Is the sample too variable? .... 193, 418
Lie Detector ............. 191, 417
three or more samples...... 307, 457
two samples in many categories ........ 245, 433
Yates correction for 2 samples in 2 categories ........ 253, 436
cluster sampling .......... 138
coefficient of determination .... 349
coefficient of multiple determination ........ 354
coincidences ............ 65
combinations ............ 103, 269
complement of an event .... 63
Concatenation does not imply causation ........ 318
conditional probability .... 77
confidence level ........ 128
contingency table .......... 245
continuous variables ........ 99
correlation coefficient ........ 346
cumulative normal frequency ........ 167
cyclical sequences ........ 193
data
four types ........ 176
interval ........ 176
nominal ........ 176
ordinal ........ 176
ratio ........ 176
data mining ........ 130
decile ........ 41
design variable ........ 358
dichotomous variable ........ 358
discrete variables ........ 99
dummy variable ........ 358
e ........ 96
Emergency Statistics Guide ........ 326
event in a sample space ........ 63
complement ........ 60
independent events ........ 58, 75
intersection of two events .... 65
mutually exclusive ........ 64
union of two events ........ 64
exponential distribution .... 100, 400
second form ........ 101
third form ........ 101
extended hypergeometric distribution ........ 110
F-Distribution test ........ 234, 428
factorial ........ 97
Field Guide ........ 388
finite population correction factor .... 157
firecracker factory ........ 29
Fisher, R. A. ........ 163
Fisher’s Exact test ........ 254, 437
frequency distribution ........ 21
Gaussian distribution .... 118, 401
Gosset, William Sealey .... 163
histogram ........ 32, 169
Hosenaufschlag Macht Geld .... 192
Hume’s problem ........ 126
hypergeometric distribution .... 106, 395
extended ........ 110, 396
induction ........ 126, 144
<table>
<thead>
<tr>
<th>Term</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>inferential statistics</td>
<td>91</td>
</tr>
<tr>
<td>interval data</td>
<td>176</td>
</tr>
<tr>
<td>Journal of Fredometrika</td>
<td>210</td>
</tr>
<tr>
<td>Kingie</td>
<td>43</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov goodness-of-fit test</td>
<td>180, 409</td>
</tr>
<tr>
<td>for a normal distribution</td>
<td>184</td>
</tr>
<tr>
<td>for a uniform distribution</td>
<td>180</td>
</tr>
<tr>
<td>Kruskal-Wallis test</td>
<td>300, 453</td>
</tr>
<tr>
<td>post-test</td>
<td>302, 456</td>
</tr>
<tr>
<td>Leibnitz Lane</td>
<td>34</td>
</tr>
<tr>
<td>leptokurtic</td>
<td>49</td>
</tr>
<tr>
<td>level of significance of a test</td>
<td>128</td>
</tr>
<tr>
<td>Lilliefors test for normality</td>
<td>164, 411</td>
</tr>
<tr>
<td>linear regression</td>
<td>27, 346</td>
</tr>
<tr>
<td>ln x</td>
<td>363</td>
</tr>
<tr>
<td>logistic regression</td>
<td>372</td>
</tr>
<tr>
<td>Luther’s Table Talk</td>
<td>459</td>
</tr>
<tr>
<td>Mann-Whitney test</td>
<td>243, 431</td>
</tr>
<tr>
<td>maximum likelihood estimators</td>
<td>387</td>
</tr>
<tr>
<td>mean average</td>
<td>31</td>
</tr>
<tr>
<td>median average</td>
<td>39</td>
</tr>
<tr>
<td>mesokurtic</td>
<td>49</td>
</tr>
<tr>
<td>MLE</td>
<td>387</td>
</tr>
<tr>
<td>mode average</td>
<td>37</td>
</tr>
<tr>
<td>Monte Carlo method</td>
<td>207</td>
</tr>
<tr>
<td>mu</td>
<td>31</td>
</tr>
<tr>
<td>multicollinearity</td>
<td>358</td>
</tr>
<tr>
<td>multinomial distribution</td>
<td>108, 398</td>
</tr>
<tr>
<td>multiple coefficient of regression</td>
<td>354</td>
</tr>
<tr>
<td>multiple regression</td>
<td>353</td>
</tr>
<tr>
<td>nominal data</td>
<td>176</td>
</tr>
<tr>
<td>nonlinear regression</td>
<td>363</td>
</tr>
<tr>
<td>nonparametric statistics</td>
<td>177</td>
</tr>
<tr>
<td>normal distribution</td>
<td>115, 401</td>
</tr>
<tr>
<td>Normal Distribution—large sample, large part of the population</td>
<td>158, 403</td>
</tr>
<tr>
<td>Normal Distribution—large sample, small part of the population</td>
<td>146, 402</td>
</tr>
<tr>
<td>normal equations</td>
<td>354</td>
</tr>
<tr>
<td>null hypothesis</td>
<td>125</td>
</tr>
<tr>
<td>Oeuf Cubique</td>
<td>278</td>
</tr>
<tr>
<td>one sample with two variables</td>
<td>251, 435</td>
</tr>
<tr>
<td>one-tail vs. two-tail</td>
<td>133, 219</td>
</tr>
<tr>
<td>One-Way ANOVA test for Independent Samples</td>
<td>270, 438</td>
</tr>
<tr>
<td>One-Way ANOVA test for Matched Samples</td>
<td>279, 441</td>
</tr>
<tr>
<td>ordinal data</td>
<td>176</td>
</tr>
<tr>
<td>outlier</td>
<td>37, 140</td>
</tr>
<tr>
<td>( \phi_{H</td>
<td>V} )</td>
</tr>
<tr>
<td>parametric statistics</td>
<td>177</td>
</tr>
<tr>
<td>past</td>
<td>388</td>
</tr>
<tr>
<td>Pearson Product Moment Correlation</td>
<td></td>
</tr>
<tr>
<td>Coefficient for Sample Data</td>
<td>348</td>
</tr>
<tr>
<td>Pearson, Karl</td>
<td>163</td>
</tr>
<tr>
<td>percent correct predictions statistic</td>
<td>387</td>
</tr>
<tr>
<td>percentile</td>
<td>41</td>
</tr>
<tr>
<td>permutations</td>
<td>102</td>
</tr>
<tr>
<td>pi (for circles)</td>
<td>30</td>
</tr>
<tr>
<td>platykurtic</td>
<td>49</td>
</tr>
<tr>
<td>point estimate</td>
<td>152</td>
</tr>
<tr>
<td>Poisson distribution</td>
<td>95, 399</td>
</tr>
<tr>
<td>populations vs. samples</td>
<td>30</td>
</tr>
<tr>
<td>Post-test for One-Way ANOVA for independent samples</td>
<td>273, 440</td>
</tr>
<tr>
<td>Post-test for One-Way ANOVA for matched samples</td>
<td>283, 444</td>
</tr>
<tr>
<td>power of a test</td>
<td>175</td>
</tr>
<tr>
<td>prediction interval</td>
<td>347, 350</td>
</tr>
<tr>
<td>present</td>
<td>347</td>
</tr>
<tr>
<td>proportion</td>
<td>30, 48, 106, 108, 110, 116</td>
</tr>
<tr>
<td>binomial distribution—large sample</td>
<td>152, 405</td>
</tr>
<tr>
<td>binomial distribution—small sample</td>
<td>204, 406</td>
</tr>
<tr>
<td>quartile</td>
<td>41</td>
</tr>
<tr>
<td>quintile</td>
<td>41</td>
</tr>
<tr>
<td>ratio data</td>
<td>176</td>
</tr>
<tr>
<td>regression line</td>
<td>347</td>
</tr>
<tr>
<td>Runs test</td>
<td>194, 419</td>
</tr>
<tr>
<td>sample space</td>
<td>57, 63</td>
</tr>
<tr>
<td>event</td>
<td>58</td>
</tr>
<tr>
<td>samples</td>
<td></td>
</tr>
<tr>
<td>blocked</td>
<td>279</td>
</tr>
<tr>
<td>determining sample size</td>
<td>148</td>
</tr>
<tr>
<td>how to take one</td>
<td>124</td>
</tr>
<tr>
<td>independent samples</td>
<td>227</td>
</tr>
</tbody>
</table>
**Index**

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>paired samples</td>
<td>216</td>
</tr>
<tr>
<td>pilot samples</td>
<td>132</td>
</tr>
<tr>
<td>sensitive questions</td>
<td>134</td>
</tr>
<tr>
<td>stratified sample</td>
<td>139</td>
</tr>
<tr>
<td>systematic sample</td>
<td>138</td>
</tr>
<tr>
<td>Ten Rules of Fair Play</td>
<td>130</td>
</tr>
<tr>
<td>Two Large Independent Samples</td>
<td>230, 426</td>
</tr>
<tr>
<td>Two Normal Independent Samples</td>
<td>232, 427</td>
</tr>
<tr>
<td>Two Normal Independent Samples with known sigmas</td>
<td>232, 427</td>
</tr>
<tr>
<td>Santa Clausing Village</td>
<td>34</td>
</tr>
<tr>
<td>saturated model</td>
<td>360</td>
</tr>
<tr>
<td>scatter diagram</td>
<td>22</td>
</tr>
<tr>
<td>sequence</td>
<td>193</td>
</tr>
<tr>
<td>random/cyclical/trend</td>
<td>193</td>
</tr>
<tr>
<td>sigma notation</td>
<td>44</td>
</tr>
<tr>
<td>Sign test</td>
<td>173, 414</td>
</tr>
<tr>
<td>Sign test for nominal data</td>
<td>177, 415</td>
</tr>
<tr>
<td>Sign test for paired samples of Hot &amp; Cold</td>
<td>226, 424</td>
</tr>
<tr>
<td>Sign test for two paired samples</td>
<td>223, 423</td>
</tr>
<tr>
<td>simple random sample</td>
<td>137</td>
</tr>
<tr>
<td>skewed curves</td>
<td>49</td>
</tr>
<tr>
<td>left</td>
<td>172</td>
</tr>
<tr>
<td>slope-intercept form</td>
<td>24</td>
</tr>
<tr>
<td>Smith-Satterthwaite test</td>
<td>239, 430</td>
</tr>
<tr>
<td>burglar's use of this test</td>
<td>240</td>
</tr>
<tr>
<td>squared multiple R</td>
<td>354</td>
</tr>
<tr>
<td>standard deviation</td>
<td>44</td>
</tr>
<tr>
<td>of a population</td>
<td>47</td>
</tr>
<tr>
<td>of a sample</td>
<td>47</td>
</tr>
<tr>
<td>standardizing the data</td>
<td>165</td>
</tr>
<tr>
<td>statistically significant</td>
<td>141</td>
</tr>
<tr>
<td>step down method</td>
<td>361</td>
</tr>
<tr>
<td>Student's t-Distribution</td>
<td>163, 404</td>
</tr>
<tr>
<td>symmetry</td>
<td>169</td>
</tr>
<tr>
<td>ten lollipops</td>
<td>204</td>
</tr>
<tr>
<td>Ten Rules of Fair Play</td>
<td>130</td>
</tr>
<tr>
<td>The Chart</td>
<td>48</td>
</tr>
<tr>
<td>topology</td>
<td>292</td>
</tr>
<tr>
<td>trends in sequences</td>
<td>193</td>
</tr>
<tr>
<td>Two Normal Independent Samples with known sigmas</td>
<td>232</td>
</tr>
<tr>
<td>two paired samples tests—four of them</td>
<td>217, 420</td>
</tr>
<tr>
<td>Two Proportions with 2 samples in 2 categories test</td>
<td>227, 425</td>
</tr>
<tr>
<td>two small independent samples</td>
<td></td>
</tr>
<tr>
<td>when standard deviations roughly equal</td>
<td>234, 429</td>
</tr>
<tr>
<td>with different standard deviations</td>
<td>239, 430</td>
</tr>
<tr>
<td>Two-Factor ANOVA</td>
<td>285, 445</td>
</tr>
<tr>
<td>post-test</td>
<td>288, 448</td>
</tr>
<tr>
<td>Two-Factor ANOVA with many observations per cell</td>
<td>294, 449</td>
</tr>
<tr>
<td>type I error</td>
<td>128</td>
</tr>
<tr>
<td>type II error</td>
<td>128</td>
</tr>
<tr>
<td>uniform distribution</td>
<td>169</td>
</tr>
<tr>
<td>unimodal</td>
<td>50</td>
</tr>
<tr>
<td>Vagrancy Case of Fred Gauss vs. the State of Kansas</td>
<td>77</td>
</tr>
<tr>
<td>variance</td>
<td>44, 192</td>
</tr>
<tr>
<td>Venn diagrams</td>
<td>63</td>
</tr>
<tr>
<td>Wald interval</td>
<td>153</td>
</tr>
<tr>
<td>weighted average</td>
<td>272</td>
</tr>
<tr>
<td>Wilcoxon Rank-Sum test</td>
<td>431</td>
</tr>
<tr>
<td>Wilcoxon Signed Ranks test</td>
<td>169, 413</td>
</tr>
<tr>
<td>Wilcoxon Signed Ranks test for two paired samples</td>
<td>220, 421</td>
</tr>
<tr>
<td>y = mx + b</td>
<td>24</td>
</tr>
<tr>
<td>y hat</td>
<td>356</td>
</tr>
<tr>
<td>z-score</td>
<td>117</td>
</tr>
<tr>
<td>standardizing the data</td>
<td>165</td>
</tr>
</tbody>
</table>