

*Life of Fred*<sup>®</sup>  
*Statistics*  
*Expanded Edition*

Stanley F. Schmidt, Ph.D.



Polka Dot Publishing

## *Statistics*

**D**ecisions! 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 campaign that 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, army, or 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 might 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 there, 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, primes, "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 Expanded Edition*. ( $\hat{y}$  is read "y-hat.")

## WHAT BACKGROUND DO I NEED?

It would be nice to have a little algebra so that  $x^2$ , 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 do any algebra word problems.\*

We'll use the greater than sign ( $>$ ) and plus-or-minus ( $\pm$ ).

---

\* None of those old word problems like: JACKIE IS CHASING DALE DOWN THE HALL WITH AN AX. JACKIE IS TRAVELING 7 FT/SEC AND DALE IS RUNNING AT 5 FT/SEC. THEY ARE 8 FEET APART. HOW SOON SHOULD DALE 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 \$8.) In a couple of years they will probably be included free in cereal boxes.

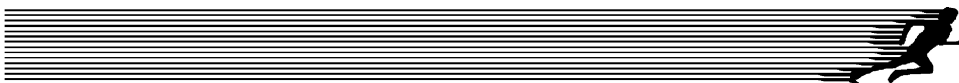
### ANY SPECIAL SUGGESTIONS BEFORE I START CHAPTER 1?

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. In this Expanded Edition, all the answers are supplied.

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 1, or to the Table of Contents, or to the Index.

Instead, turn to the Emergency Statistics Guide which begins on page 353. 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.

## *A Note to Teachers*

Sometimes life suddenly gets a lot easier. *Life of Fred: Statistics Expanded Edition* might be the best teaching assistant that you've ever had. For your students this book will be much more than just a source of homework. Open this book at random and you will see why many of your students will actually read this textbook. (Gasp! Shock!) That will make your job significantly easier.

This book has lots and lots of statistics. More than enough for most classroom courses. How many other beginning statistics books teach all of:

Kolmogorov-Smirnov  
Fisher's Exact Test  
 $\chi^2$  with Yates correction  
Smith-Satterthwaite Test  
Two-Factor ANOVA with many observations per cell  
Agresti-Coull confidence intervals

along with all of the more familiar topics (finding the mean average, finding the standard deviation, drawing a histogram, etc.) that every textbook has? With all these topics in this book, you have plenty of flexibility to include just those items that you most enjoy teaching. Just take a peek at the Table of Contents on page 15.

Your students will appreciate the fact that you have chosen a textbook that:

- . . . is fun to read,
- . . . costs about one-third of other beginning statistics books,
- . . . has a Field Guide, starting on page 420, which brings each test into sharp focus,
- . . . includes an Emergency Statistics Guide, starting on page 353, which (almost) instantly can direct you from any problem to the right test to use.

This is not one of those textbooks that tells you to "Turn to Example 18-6" and doesn't tell you what page it's on. Or Table 18-6. Or Problem 18-6. Or Equation 18-6. Or Figure 18-6. It can be maddening to have to turn back through 20 pages looking for that example/table/problem/equation/figure. That won't happen in this book. We use page numbers.

## Some Ideas on Teaching with *LOF: Statistics*

1. Each chapter has several *Your Turn to Play* sections with representative problems and their complete solutions. Many teachers find this the ideal place to begin their discussions.
2. At the end of each chapter are six sets of problems (called Cities). Each City may take your students 20–40 minutes to work through. The answers are all supplied, but not the complete solutions. With the answers given, students will know whether they have done the problem correctly. Without the solutions supplied in the text, you will be able to tell whether the students have actually worked the problem.

Why *Cities*? This makes it easier on you if all you have to say is, “Do San Francisco for homework” rather than the old, “Do every third problem on page 231.”
3. It is expected that each student will work through all the *Your Turn to Play* sections and all of the Cities problems.
4. Ask your students to read the material the night before you cover it in class. *The nature of this book makes that kind of assignment possible.* That will make your teaching of the material much more pleasant.
5. The heart of the book—at least for me—is “The Art of the Sample,” which is Chapter 4½ that begins on page 133. Students who master all 46 statistical tests in this book will have a powerful arsenal at their disposal which can be used to spread truth or to deceive. The Ten Rules of Fair Play as described in Chapter 4½ set out some ethical guidelines for the use of that arsenal of tests. This can promote some very interesting classroom discussions. Some students love to find the “gray areas” in any set of rules they’re expected to follow. Rule #1, for example, prohibits data mining. A discussion might revolve around, “Is it really data mining if you happen to notice that many politicians seem to cheat on their income taxes and then you do a survey on that topic?”



# *Contents*

Chapter 1	Descriptive Statistics. . . . . 21
	<ul style="list-style-type: none"> <li>frequency distributions</li> <li>scatter diagrams</li> <li>averages—mean, median, and mode</li> <li>linear regression</li> <li>populations vs. samples</li> <li>histograms</li> <li>range</li> <li>percentiles, deciles, quintiles, quartiles</li> <li>variance</li> <li>sigma notation</li> <li>standard deviation for populations and for samples</li> <li>distributions—skewed, platykurtic, leptokurtic, bimodal</li> </ul>
Chapter 2	Probability. . . . . 59
	<ul style="list-style-type: none"> <li>outcomes</li> <li>sample space</li> <li>events—-independent, complements, mutually exclusive</li> <li>Venn diagrams</li> </ul>
Chapter 3	Conditional Probability. . . . . 77
	<ul style="list-style-type: none"> <li><math>\mathcal{P}(A   B)</math> notation</li> <li>definition of conditional probability</li> <li>Bayes' theorem and its proof</li> <li>generalized Bayes' theorem</li> </ul>
Chapter 3½	Looking Forward to the Next Four Chapters. . . . . 97
	<ul style="list-style-type: none"> <li>the Future—zero samples</li> <li>the Past—one sample</li> <li>the Present—two samples</li> <li>the Present—three or more samples</li> </ul>
Chapter 4	The Future—Zero Samples. . . . . 101
	<ul style="list-style-type: none"> <li>Poisson distributions</li> <li>e</li> <li>factorial</li> <li>continuous vs. discrete variables</li> <li>exponential distributions—three forms</li> <li>permutations and combinations</li> <li>Bernoulli variables</li> <li>binomial distributions</li> <li>hypergeometric distributions</li> <li>multinomial distributions</li> </ul>

extended hypergeometric distributions  
normal distributions—Gaussian distributions  
normal curves to approximate binomial distributions

## Chapter 4½    The Art of the Sample. . . . . 133

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. . . . . 155

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. . . . .	217
	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 <i>The Journal of Fredometrika</i> ),	
	which describes a new approach to the problem	
Chapter 6	The Present—Two Samples. . . . .	227
	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	
	<i>F</i> -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. . . . .	283
	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. . . . .	353
Chapter 8	Finding Regression Equations. . . . .	369
	linear regression	
	prediction intervals	
	Pearson Product Moment Correlation Coefficient (r)	
	coefficient of determination	
	multiple regression	
	normal equations	
	coefficient of multiple determination ( $R^2$ )	
	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. . . . . 420

**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
- 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,  $n \leq 30$ )
- 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 ( $\mu_1 - \mu_2$ )
- 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,  $n \geq 30$
- Two Independent Samples ( $\sigma_1$  and  $\sigma_2$  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 ( $\chi^2$ ), two samples of nominal data in multiple categories.

One Sample with Two Variables  
 Chi-squared test ( $\chi^2$ )—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..... 490

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

Index. .... 571

## Chapter One

### Descriptive Statistics

**T**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:

But that looked much too "numberty" for Fred's taste. He liked to keep things simple. Instead of

time	no. of drops
4:13	4
4:14	6

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:

time	no. of drops
1	4
2	6

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.

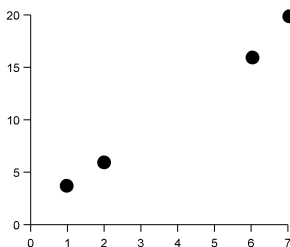
Fred had a very new machine (it was a gift from his students), but the university had very old phone lines. “ISP IS NOT RESPONDING” appeared on his screen. “ERROR 397 THE NUMBER IS BEING REDIALED.”

Fred went back to looking at the pot. It was 4:18 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:

time	no. of drops
1	4
2	6
6	16
7	20

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



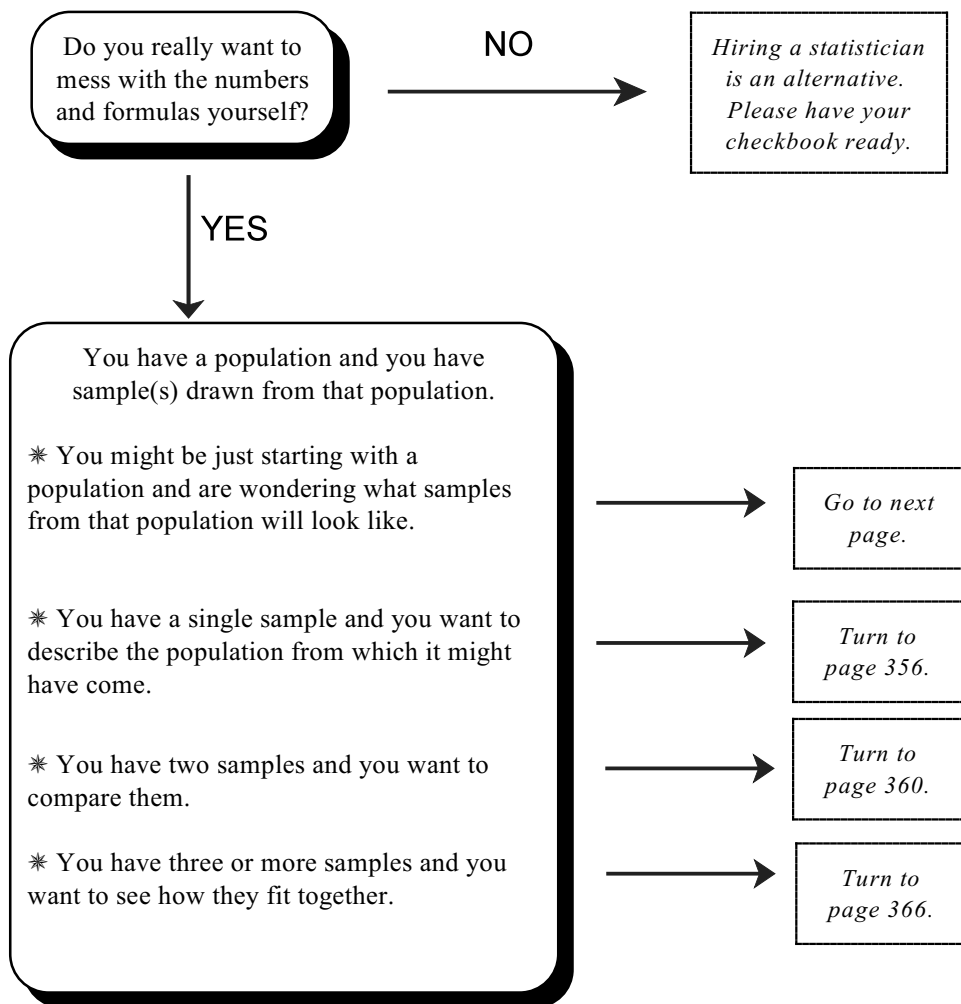
Fred's Scatter Diagram

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.

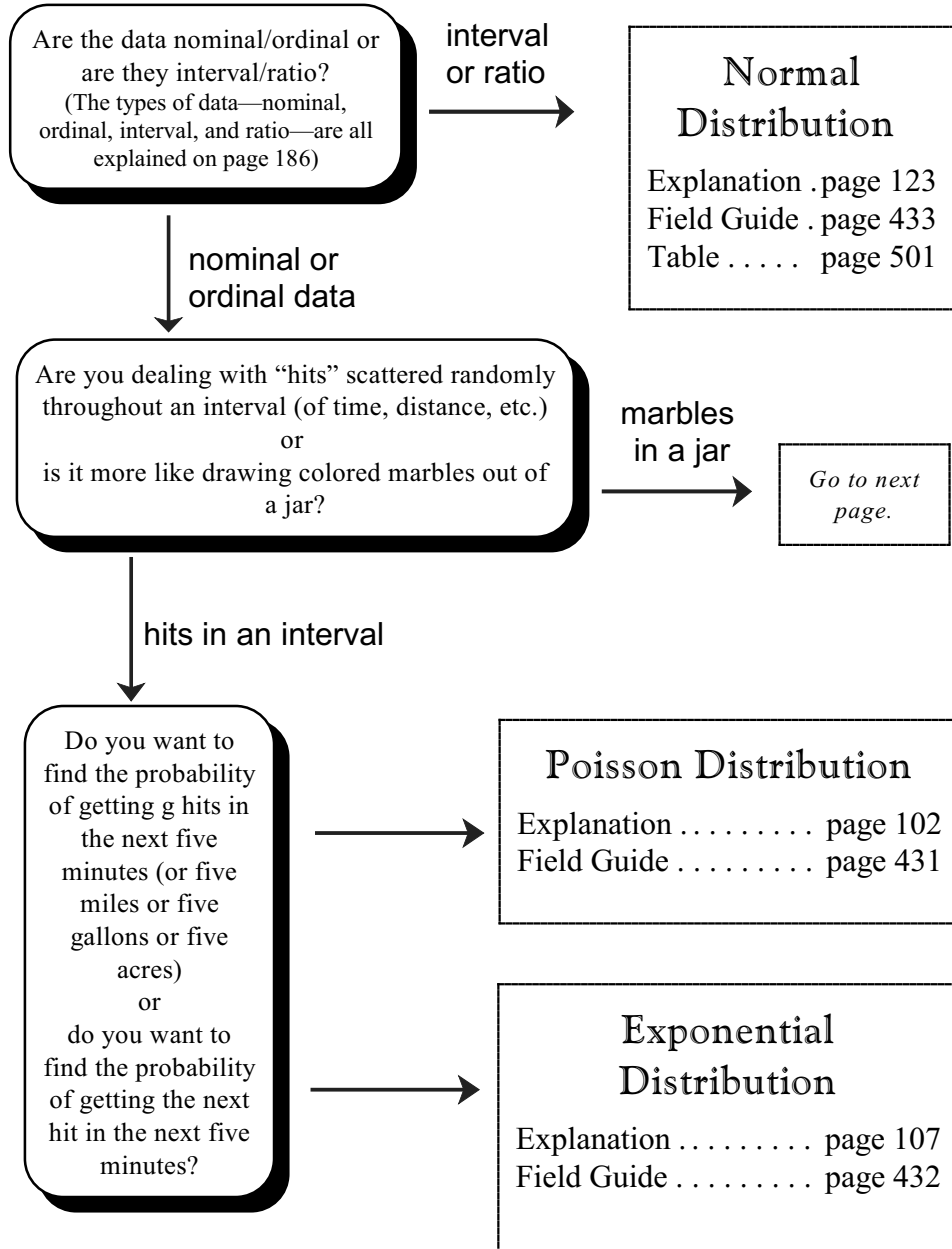


*From the previous page.*

You have a population and you want to know what samples from this population *will* look like.

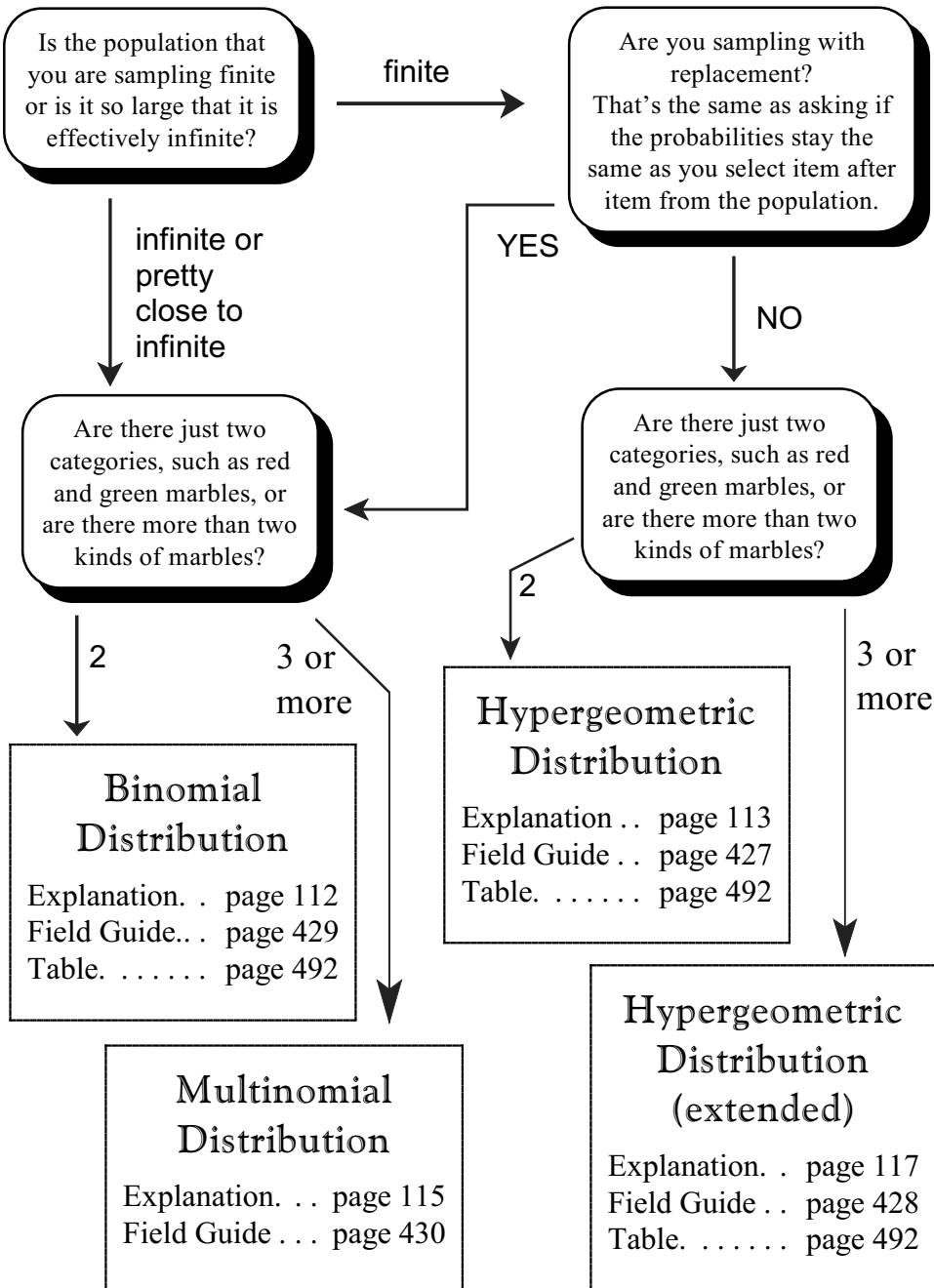
We call this the Future.

You are starting with zero samples.



*From the previous page.*

You are drawing colored marbles out of a jar.

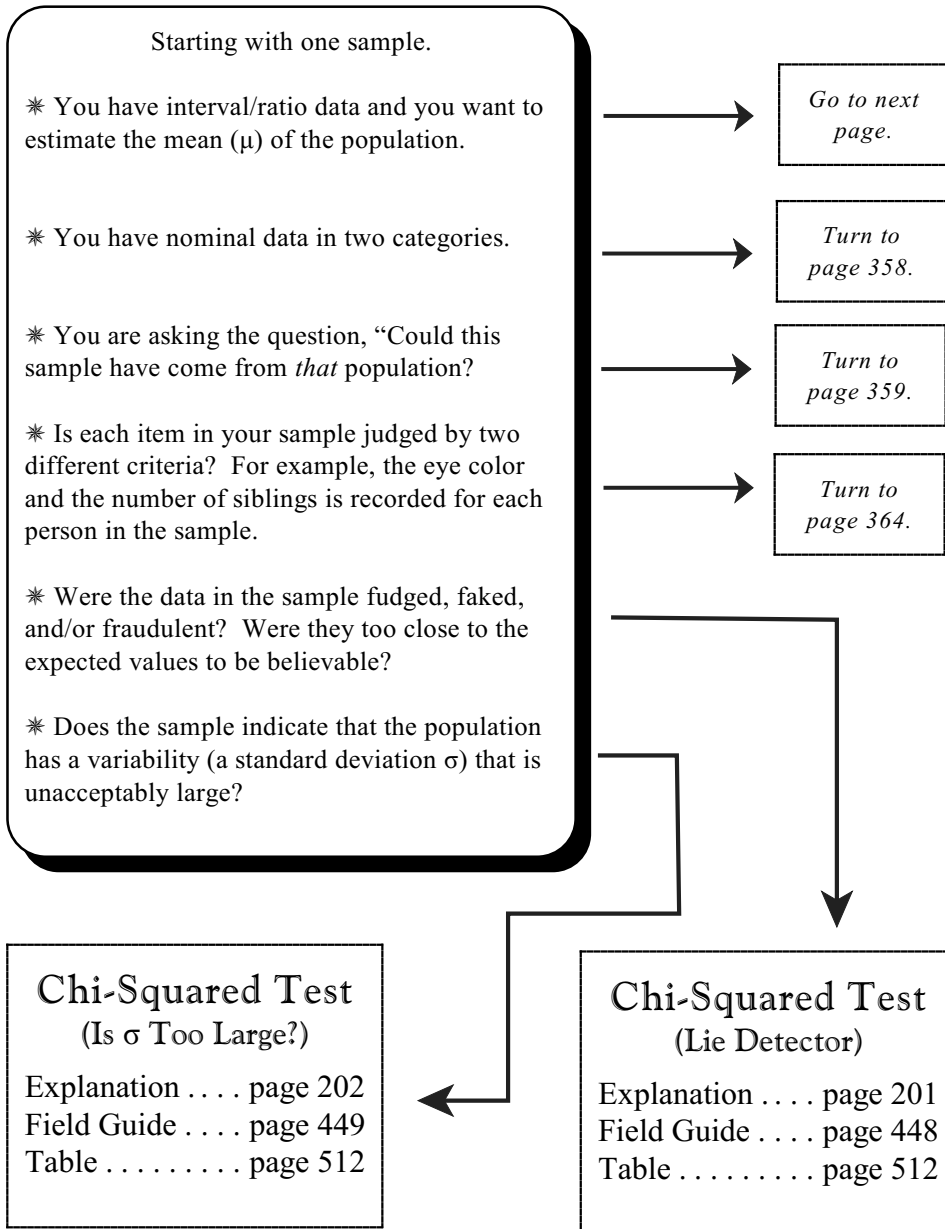


From page 353.

A single sample is known.

You want to describe the population from which it *came*.

We call this the Past.



# *Index*

adjusted coefficient of multiple determination. . . . .	381	coefficient of determination . . . . .	376
Agresti-Coull confidence interval . . . . .	163	coefficient of multiple determination. . . . .	381
ANOVA table. . . . .	308	coincidences. . . . .	68
antilog. . . . .	396	combinations. . . . .	110, 286
average		complement of an event. . . . .	66
mean. . . . .	31	correlation does not imply causation. . . . .	338
median. . . . .	39	conditional probability. . . . .	82
mode. . . . .	37	confidence level. . . . .	137
bad luck. . . . .	145	contingency table. . . . .	258
Bayes' theorem. . . . .	87	continuous variables. . . . .	106
generalized. . . . .	91	correlation coefficient. . . . .	373
proof. . . . .	88	cumulative normal frequency . . . . .	177
bell-shaped curve. . . . .	49	cyclical sequences. . . . .	203
Bernoulli variable. . . . .		data	
. . . . .	111, 217, 400	four types. . . . .	186
bimodal. . . . .	49	interval. . . . .	186
binomial distribution. . . . .	112, 429	nominal. . . . .	186
for proportions. . . . .	162	ordinal. . . . .	186
small sample. . . . .	217	ratio. . . . .	186
Central Limit Theorem. . . . .	160	data mining. . . . .	139
Chi-squared test		decile. . . . .	41
combining categories. . . . .	199	design variable. . . . .	385
goodness-of-fit test. . . . .		dichotomous variable. . . . .	385
. . . . .	197, 447	discrete variables. . . . .	106
Is the sample too variable? . . . . .	203, 449	dummy variable. . . . .	385
Lie Detector. . . . .	201, 448	e. . . . .	103
three or more samples. . . . .		Emergency Statistics Guide . . . . .	353
. . . . .	327, 488	event in a sample space. . . . .	66
two samples in many categories. . . . .	258, 464	complement. . . . .	63
Yates correction for 2 samples in 2 categories. . . . .	266, 467	independent events. . . . .	61, 80
cluster sampling. . . . .	147		

## *Index*

- intersection of two events
  - ..... 68
  - mutually exclusive. .... 67
  - union of two events..... 67
- exponential distribution. ....
  - ..... 107, 432
  - second form. .... 108
  - third form. .... 108
- extended hypergeometric
  - distribution..... 117
- F*-distribution test .... 247, 459
- factorial. .... 104
- Field Guide..... 420
- finite population correction factor
  - ..... 167
- firecracker factory. .... 29
- Fisher, R. A..... 173
- Fisher's Exact test. .... 267, 468
- frequency distribution. .... 21
- Gaussian distribution. . . 125, 433
- Gosset, William Sealey. .... 173
- histogram. .... 32, 179
- Hosenaufschlag Macht Geld*
  - ..... 202
- Hume's problem.. .... 135
- hypergeometric distribution
  - ..... 113, 427
  - extended. .... 117, 428
- induction. .... 135, 153
- inferential statistics. .... 97
- interval data. .... 186
- Journal of Fredometrika..... 223
- Kingie..... 43
- Kolmogorov-Smirnov goodness-
  - of-fit test..... 190, 439
  - for a normal distribution
    - ..... 194
  - for a uniform distribution
    - ..... 190
- Kruskal-Wallis test .... 319, 484
  - post test..... 321, 487
- Leibnitz Lane..... 34
- leptokurtic. .... 49
- level of significance of a test
  - ..... 137
- Lilliefors test for normality. ....
  - ..... 174, 442
- linear regression..... 27, 373
- ln x. .... 390
- logistic regression. .... 399
- Luther's *Table Talk*. .... 490
- Mann-Whitney test.... 256, 462
- maximum likelihood estimators
  - ..... 419
- mean average. .... 31
- median average. .... 39
- mesokurtic. .... 49
- MLE..... 419
- mode average..... 37
- Monte Carlo method. .... 220
- mu ( $\mu$ )..... 31
- multicollinearity..... 385
- multinomial distribution. ....
  - ..... 115, 430
- multiple coefficient of regression
  - ..... 381
- multiple regression..... 380
- nominal data..... 186
- nonlinear regression..... 390
- nonparametric statistics. .... 187
- normal distribution.... 122, 433
- Normal Distribution—large
  - sample, large part of the
    - population. .... 168, 435
- Normal Distribution—large
  - sample, small part of the
    - population. .... 156, 434
- normal equations. .... 381

## *Index*

- null hypothesis. . . . . 134
- Oeuf Cubique*. . . . . 295
- one sample with two variables
  - . . . . . 264, 466
- one-tail vs. two-tail. . . . 142, 232
- One-Way ANOVA test for
  - independent samples
    - . . . . . 287, 469
- One-Way ANOVA test for
  - matched samples. . . . .
    - . . . . . 297, 472
- ordinal data. . . . . 186
- outlier. . . . . 37, 149
- $P(H | V)$ . . . . . 79
- parametric statistics. . . . . 187
- past. . . . . 420
- Pearson Product Moment
  - Correlation Coefficient
    - for Sample Data. . . . . 375
- Pearson, Karl. . . . . 173
- percent correct predictions
  - statistic. . . . . 419
- percentile. . . . . 41
- permutations. . . . . 109
- $\pi$  ( $\pi$ ). . . . . 30
- platykurtic. . . . . 49
- point estimate. . . . . 162
- Poisson distribution. . . . 102, 431
- populations vs. samples. . . . 30
- Post Test for One-Way ANOVA
  - for independent samples
    - . . . . . 290, 471
- Post Test for One-Way ANOVA
  - for matched samples
    - . . . . . 301, 475
- power of a test. . . . . 185
- prediction interval. . . . . 374, 377
- present. . . . . 420
- proportion. . . . . 30, 48, 113, 115,
  - 117, 123
  - binomial distribution—large
    - sample. . . . . 162, 437
  - binomial distribution—small
    - sample. . . . . 217, 438
- quartile. . . . . 41
- quintile. . . . . 41
- ratio data. . . . . 186
- regression line. . . . . 374
- Runs test . . . . . 204, 450
- sample space. . . . . 60, 66
  - event. . . . . 61
- samples
  - blocked. . . . . 297
  - determining sample size
    - . . . . . 158
  - how to take one. . . . . 133
  - independent samples. . . . . 240
  - paired samples. . . . . 229
  - pilot samples. . . . . 141
  - sensitive questions. . . . . 143
  - stratified sample. . . . . 148
  - systematic sample. . . . . 147
  - Ten Rules of Fair Play. . . . 139
  - Two Large Independent
    - Samples. . . . . 243, 457
  - Two Normal Independent
    - Samples with known
      - sigmas. . . . . 245, 458
- Santa Clausing Village. . . . . 34
- saturated model. . . . . 387
- scatter diagram. . . . . 22
- sequence
  - random/cyclical/trend. . . . 203
- sigma notation. . . . . 44
- Sign test . . . . . 183, 445
- Sign test for nominal data . . . .
  - . . . . . 187, 446

Sign test for paired samples of  
     Hot & Cold . . . . 239, 455  
 Sign test for two paired samples  
     . . . . . 236, 454  
 simple random sample. . . . . 146  
 skewed curves. . . . . 49  
     left. . . . . 182  
 slope-intercept form. . . . . 24  
 Smith-Satterthwaite test . . . . .  
     . . . . . 252, 461  
     burglar's use of this test. . . 253  
 squared multiple R. . . . . 381  
 standard deviation. . . . . 44  
     of a population ( $\sigma$ ). . . . . 47  
     of a sample ( $s$ ). . . . . 47  
 standardizing the data. . . . . 175  
 statistically significant. . . . . 150  
 step down method. . . . . 388  
 Student's  $t$ -distribution. . . . .  
     . . . . . 173, 436  
 symmetry. . . . . 179  
 ten lollipops. . . . . 217  
 Ten Rules of Fair Play. . . . . 139  
 The Chart. . . . . 48  
 topology. . . . . 310  
 trends in sequences. . . . . 203  
 Two Normal Independent  
     Samples with known  
     sigmas. . . . . 245  
 two paired samples tests—four of  
     them. . . . . 230, 451  
 Two Proportions with 2 samples  
     in 2 categories test. . . . .  
     . . . . . 240, 456  
 two small independent samples  
     when standard deviations  
     roughly equal. . . 247, 460  
     with different standard  
     deviations. . . . . 252, 461  
 Two-Factor ANOVA. . . 303, 476  
     post test. . . . . 306, 479

Two-Factor ANOVA with many  
     observations per cell  
     . . . . . 313, 480  
 type I error. . . . . 137  
 type II error. . . . . 137  
 uniform distribution. . . . . 179  
 unimodal. . . . . 50  
 Vagrancy Case of Fred Gauss vs.  
     the State of Kansas. . . . 82  
 variance. . . . . 44, 202  
 Venn diagrams. . . . . 66  
 Wald interval. . . . . 163  
 weighted average. . . . . 289  
 Wilcoxon Rank-Sum test . . . 462  
 Wilcoxon Signed Ranks test  
     . . . . . 179, 444  
 Wilcoxon Signed Ranks test for  
     two paired samples. . . . .  
     . . . . . 233, 452  
 $y = mx + b$ . . . . . 24  
 $\hat{y}$ . . . . . 383  
 z-score. . . . . 124  
     standardizing the data. . . . 175