Random Walks and the Binomial Distribution

Many statistics books start their treatment of inference by looking at proportions. And when you do proportions, if you want to understand what's going on at the deepest level, you have to understand how binomial distributions work.

We aren't going to use combinatorics here; you won't see a factorial in the place. Instead, we will look at some phenomena associated with the binomial distribution empirically rather than theoretically.

In this section, you will find:

"Flipping Coins—the Law of Large Numbers"
With this demo, you can flip lots of coins and see how the proportion of heads approaches one-half. Then you can do it over and over, and see the myriad paths that proportion takes, always to the same destination.

"How Random Walks Go as Root *N*"
If you take a random walk, on the average you'll wind up where you started. But the average distance you will be from the starting place increases as you keep walking; in fact, it increases without bound. Specifically, it is proportional to the square root of the number of steps. We'll see that happen in Fathom.

"Building the Binomial Distribution"
In this demo, you build the binomial distribution from the ground up—by sampling—and compare distributions with different population proportions.

"More Binomial"
Here we look at the binomial distribution slightly differently; in the previous demo, we looked at the distribution of proportions, but here we look at the distribution of counts as well. We also see how both change with sample size. You will see surprisingly subtle things; understanding them will help your understanding of the whole picture.

"Two-Dimensional Random Walks"
What would happen if you walked randomly in two dimensions? We explore that here in Fathom. You'll see bizarre results for small numbers of steps that, like so much else in statistics, look smooth and normal with large numbers.

Remember: **Sol** means that there's a solution at the back of the book.

Measures of Center and Spread

What does *mean* really mean? How can we get a feel for standard deviation? These are some of the basics, and without these topics, sampling distributions and confidence intervals won't make much sense.

The demos in this section address some interesting properties of measures of center and spread, as well as giving you some of the background that a dynamic Fathom document can do so well.

Why pay attention to this kind of thing? Even if you're an expert in center-and-spread, you will see throughout the book that, especially with small samples, the *spread* of values plays an unexpectedly powerful role. We're used to the measures of *center* as the most relevant summary—but it just isn't always the case.

In this section, you will find:

"The Meaning of Mean"
Here you see what happens to the mean when you drag a point. The mean moves, but how far?

"Mean and Median"
It's easy to say that the median is a resistant measure. But what does that really mean?

"What Do Normal Data Look Like?"
Here you look at some random numbers pulled from a normal distribution. You control the mean, standard deviation, and the size of the sample. The large sample shows the characteristic bell curve—but the small sample does not.

"Transforming the Mean and Standard Deviation"
When you add a constant to every data value, what happens? You change the mean, but not the standard deviation. What if you multiply by a constant? In this demo, you get to see it all happen dynamically; you'll be better able to "feel" the effect of these basic transformations after watching it happen.

"The Mean Is Least Squares, Too"
This demo, about the meaning of mean, is perhaps best approached when studying more advanced topics (such as the sample variance). Basically, the demo shows how you can define the mean as the number that minimizes the sum of the squares of the deviations from the data points. And what if you used absolute deviation instead of squares? Then you would get the *median*.

Remember: **Sol** means that there's a solution at the back of the book.

Regression and Correlation

This section is about the relationship between two continuous variables. When you use one to predict the other, you can use—among other things—a least-squares regression line. When you want to describe how two variables are related, you can use the correlation coefficient. Regression lines often show up in the "modeling" section of a statistics course, and they belong there. But these techniques appear throughout inferential statistics too, as we shall see.

These topics can be tricky; the demos in this section will help clarify a few of the most troubling issues.

In this section, you will find:

"Least-Squares Linear Regression"
What are the *squares* in least-squares regression? In Fathom, you can show them in this simple but essential demo.

"Standard Scores"
Just as we use proportions to compare two groups of different sizes, we can use standard scores to compare—or, in this case, combine—data from different distributions. The standard score is a measure of where something is in relation to the population, as measured in terms of standard deviation. This "dimensionlessness" is crucial to its success—and we'll use the same principle to create the t statistic a little later.

"Devising the Correlation Coefficient"
Bit by bit, we construct a statistic that measures the strength of a relationship, using standard scores as a starting point. The elegance of cross-multiplying may inspire you to be creative in the way you design your own statistics.

"Correlation Coefficients of Samples"
In a way, this is skipping ahead, giving you an early look at another big theme in these demos: sampling distributions. We will see that a sample may have a very different correlation than its source population. Things can seem correlated when they aren't, and conversely.

"Regression Toward the Mean"
Why is it that the people who do best seem to fall the farthest? Fathom can show you. This phenomenon also helps us see how the regression line and the correlation coefficient measure different things.

Remember: **Sol** means that there's a solution at the back of the book.

Standard Deviation, Standard Error, and Student's *t*

We already talked about standard deviation in the opening set of demos in this book, "Measures of Center and Spread." And in Demo 7, "Standard Scores," we used that spread to create the z-score—a way of describing a point in the context of its distribution, independent of how we measure the data.

This section takes these ideas further, using another, related measure of spread: standard error.

If you're learning statistics—for the first or for the nth time—it may not be clear what standard error is and why you should care. Similarly, it's easy to treat the t-statistic only as something that the computer makes for you in order to feed the black box of the t-test.

These demos give you one way to make sense of these slippery concepts.

In this section, you will find:

"Standard Error and Standard Deviation"
This demo lets you experience these two statistics visually and see how they change with sample size.

"What Is Standard Error, Really?"
Here we look at the standard error as the spread in the sampling distribution of the mean. We also look at that spread more quantitatively and, as we did in Demo 12, "How Random Walks Go as Root N," see a root-N dependence.

"The Road to Student's *t*"
Now we add another layer of complexity: Suppose we have a sample and want to figure out how far the sample's mean might be from that of the population. We discover that, *if we measure that distance in terms of the standard error,* the distribution we get is not normal. We need a new distribution: t.

"A Close Look at the *t*-Statistic"
Here we use a fictitious sample of three points and see what happens to the t-statistic as we drag the points around. We see how we can get a big t if we put all the points on one side (that's obvious—their mean should be far from the population value) and clump them together (that's not obvious, but it lowers the standard error, raising t).

Remember: **Sol** means that there's a solution at the back of the book.

Sampling Distributions

Sampling distributions show up often in statistics, and the idea of sampling distributions occurs throughout this book. Many of the relevant demos are in other places; but here, in this section, is a collection of additional ideas held together by this common thread.

In this section, you will find:

"The Distribution of Sample Proportions"

Here you get to see how this distribution depends on the population proportion and the sample size. Of course, we don't see only a *theoretical* (in this case, binomial) distribution—we see real samples, with all their attendant variability.

"Sampling Distributions and Sample Size"

Now we again take sampling distributions of the mean, but this time we compare these distributions for different sample sizes. As we discovered in Demo 17, "What Is Standard Error, Really?" the distribution gets skinnier as sample size goes up.

"How the Width of the Sampling Distribution Depends on *N*"

Here we look quantitatively at the spread of the sampling distribution. We use the IQR and discover the inverse root-N dependence related to the one we discovered in Demo 12, "How Random Walks Go as Root N."

"Does $n - 1$ Really Work in the SD?"

We can use sampling distribution—in this case, of the sample standard deviation—to assess whether a statistic is an unbiased estimator. We discover that it is not. See also "Sample Variance: Why the Denominator Is $n - 1$" in Appendix B.

"The Central Limit Theorem"

You change the population and see that the sampling distributions of most—but not all—statistics are usually—but not always—pretty normal.

Peripheral Issues

We have also included three demos in this section that aren't really part of the direct inferential road, but have more to do with making good estimates:

"Adding Uniform Random Variables"

If you sample from two uniform distributions and add, what's the distribution of the result? (Triangular.)

"How Errors Add"

Suppose the two things you add are measurements with a normally distributed error. What's the error in the sum? We see that we get the Pythagorean sum of the errors.

"German Tanks"

In the context of this historical problem, we look at estimators and their sampling distributions in order to see if they're unbiased or not.

Remember: **Sol** means that there's a solution at the back of the book.

Confidence Intervals

One of the main tasks in statistics is to make estimates of parameters. Calling them estimates is important: we will never know the parameters exactly, but we can know something about them. Confidence intervals (CIs) express what we know. To calculate confidence intervals is straightforward; to know what they really mean, however, is deep and tricky.

CI of a Proportion

While all confidence intervals share the same basic meaning, books often treat the confidence interval for a proportion separately. So do we:

"The Confidence Interval of a Proportion"

Here we explore one way of defining confidence interval: as the range of population values for which the observed value is *plausible*.

"Capturing with Confidence Intervals"

In this demo, we see the consequences of the more orthodox definition of the CI: If you construct intervals repeatedly, some of them miss the true value. The fraction that miss depends on the confidence level.

"Where Does That Root ($p(1 - p)$) Come From?"

That algebraic snippet is in the orthodox CI formula. Where does it come from? It's the standard deviation of yes/no data mapped onto one and zero.

"Why $np > 10$ Is a Good Rule of Thumb"

They give you these rules, like "Use the normal approximation if $np > 10$." This demo shows why that makes sense.

CI of Other Things

Mostly, this is the confidence interval of the *mean*. Some of these parallel their cousins in the proportion section:

"How the Width of the CI Depends on *N*"

This parallels Demo 24, "How the Width of the Sampling Distribution Depends on *N*." We find empirically that it's inversely proportional to the square root of the sample size.

"Using the Bootstrap to Estimate a Parameter"

If you know nothing about the population distribution, you can still estimate a parameter (a median in this case) using the sample you have. This shows the power of resampling.

"Exploring the Confidence Interval of the Mean"

If you have only three points, what does the CI look like? We drag the points to find out.

"Capturing the Mean with Confidence Intervals"

This is parallel to Demo 29, "Capturing with Confidence Intervals," but with the mean instead of the proportion. We see how, sometimes, you miss.

Remember: **Sol** means that there's a solution at the back of the book.

Hypothesis Tests

Hypothesis tests and estimates (that is, confidence intervals) constitute the backbone of inferential statistics. In Fathom, we can show the meaning of the hypothesis test through simulation. First we simulate the null hypothesis. Then we create a sampling distribution of a statistic where the null hypothesis is true. Finally, we compare our test statistic to that distribution.

Alternatively, we can use canned tests—such as the *t*-test of mean—and see dynamically how changing the data changes the results of the test. We can also test repeatedly, using the results as data so that we can study the behavior of the test as a whole.

This section includes:

"Fair and Unfair Dice"
How can we tell if a die is unfair? We get a simulated, loaded die. First we devise a statistic of unfairness. Then we create the sampling distribution of that statistic for *fair* dice. Comparing the two, we can assess how likely it is that a fair die would behave like our loaded one.

"Scrambling to Compare Means"
Here we have fake plant data, using two different fertilizers. Is the new one better? We devise a statistic, use scrambling (a randomization technique) to create a sampling distribution, and compare. This is another example of the basic meaning of a hypothesis test.

"Using a *t*-Test to Compare Means"
We explored the *t*-statistic in the section "Standard Deviation, Standard Error, and Student's *t*." Here we use the test instead of scrambling to compare the same two groups we compared in the previous demo. We will also see how changing the data changes the results.

"Another Look at a *t*-Test"
Here we look at repeated *t*-test results as data, and see how sample mean and standard deviation relate to the *P*-value of the test. We also see how random variation produces Type I errors.

"On the Equivalence of Tests and Estimates"
This quick demo shows how these hypothesis tests give some of the same information as confidence intervals. If the significance levels are equivalent, rejecting the hypothesis is the same as being outside the CI.

"Paired Versus Unpaired"
When we have repeated measures, is it better to do a paired test or an unpaired test? Paired is better, and this demo shows why.

"Analysis of Variance"
We return to the basic meaning of the hypothesis test for this one, first devising a statistic to tell—for more than two groups—if a measurement is independent of group membership, then creating the sampling distribution in the case of the null hypothesis for comparison.

Remember: **Sol** means that there's a solution at the back of the book.

Power in Tests

The null hypothesis is dull; if it's true, nothing is happening. Usually, we're trying to reject the null hypothesis—unless, of course, the null is true.

Yet the null hypothesis is hardly ever exactly true. Even if we fail to reject the null hypothesis (for example, if P is large), that doesn't mean the null *is* true—only that our results would be plausible if it *were* true. Are the two populations *exactly* identical? No. We just may not be able to show that convincingly.

So, given that we believe the null hypothesis is false, what's the chance that we get a significant result from our test? That's what we mean by *power*. And power depends on many factors. One is the sample size: The bigger the sample, the better chance you have of rejecting that null. Another is the test itself: Some tests are more powerful than others. Another is the alpha level of the test: There is a tradeoff between false positives and false negatives. And, of course, another factor is reality: How far is the actual population mean from the hypothesized value? A long way? Then you have a better chance.

This section includes:

"The Distribution of *P*-Values"

If you resample many times, and collect the distribution of P-values from the tests you do, what does that distribution look like? If the null hypothesis is true, that distribution is *flat*.

"Power"

Power—the chance of rejecting the null—is a function. In this demo, we perform a lot of tests with different true population values. That way, we construct power from empirical probabilities—as a function of the true population value. Sure enough, as we get farther from the null value, the probability approaches 1.

"Power and Sample Size"

This is the same setup as the previous demo, except we change the sample size instead. This is one way people often use power: If you think some quantity has a mean value of 6 and not 5, you can calculate the sample size you need to prove it.

"Heteroscedasticity and Its Consequences"

Inference for regression assumes that the variance of the dependent variable is the same everywhere, that is, it's *homoscedastic*. This demo shows what happens when it's not.

Remember: **Sol** means that there's a solution at the back of the book.

Distributions

In the course of these demos, we have come across several distributions—the uniform distribution, the normal, the binomial, and Student's *t*. Of course, there are many other distributions we have not seen; in this thin book we can't cover everything. But let's take a look at a few of these "other" distributions and get a sense of what they might be useful for. Most importantly, we will see that many interesting things are not normally distributed, or even close.

This section includes:

"Wait Time and the Geometric Distribution"

How many rolls does it take to get a six? How many deals will I have to wait through until I get a blackjack? With a fair die and an infinite deck, these are distributed *geometrically*. This distribution appears in a wide variety of situations and has wonderful properties.

"The Exponential and Poisson Distributions"

The continuous analog to the geometrical distribution is the *exponential:* How long will I have to wait for the next phone call, raindrop, or nuclear decay? (That's also the distribution of times *between* events.) And then, if you ask, "How many events will I have in the next ten minutes?" those numbers will be distributed according to the *Poisson*.

"Sampling Without Replacement and the Hypergeometric Distribution"

You know how we always draw cards from infinite decks. But suppose it's not infinite. Or suppose your sample size is a significant fraction of your population. Then you need the *hypergeometric* distribution. Without it, card counting doesn't work.

"The Bizarre Cauchy Distribution"

For our last demo, we'll look at something really pathological: a distribution that has no standard deviation. The samples have one, but the distribution does not. How can that be?

Remember: **Sol** means that there's a solution at the back of the book.