Fathoms
Statistics Demonstrations for Deeper Understanding

Tim Erickson
# Contents

## What Are These Demos About? xi

How to Use These Demos xii

## If This Is Your First Time Using Fathom xv

Tutorial: An Extended Example xv

A Few Good Skills xix

Fathom Overview xxi

## Measures of Center and Spread

Demo 1: The Meaning of Mean 2

The mean • How individual values affect the mean

Demo 2: Mean and Median 4

Measures of center: mean, median, and midrange • Resistance: what happens to the measures when you move one point

Demo 3: What Do Normal Data Look Like? 6

Normally distributed data • The effect of changing the mean and standard deviation

Demo 4: Transforming the Mean and Standard Deviation 8

What happens to the mean and standard deviation when you add a constant to every value or multiply every value by a constant

Demo 5: The Mean Is Least Squares, Too 10

Defining the mean as the place where the sum of squares of deviations is a minimum (just like the least-squares line) • The median, and what it minimizes

## Regression and Correlation

Demo 6: Least-Squares Linear Regression 14

Exploring the squares in least squares • Minimizing the areas of the squares built on residuals

Demo 7: Standard Scores 16

Using standard scores to compare unlike scales • Making a scale in terms of standard deviations

Demo 8: Devising the Correlation Coefficient 18

How the correlation coefficient measures what it does

© 2014 William Finzer
Demo 9: Correlation Coefficients of Samples
How samples from a correlated population yield different values for the correlation • How sample size affects that sampling distribution

Demo 10: Regression Toward the Mean
Regression toward the mean • The meaning—and asymmetry—of the least-squares line

Random Walks and the Binomial Distribution

Demo 11: Flipping Coins—the Law of Large Numbers
How the proportion of heads approaches 0.5 as sample size increases • How the number of heads does not approach half the sample size

Demo 12: How Random Walks Go as Root N
How the distance from the origin increases with the number of steps

Demo 13: Building the Binomial Distribution
Constructing the binomial distribution by resampling • How the distribution depends on the population proportion

Demo 14: More Binomial
How the binomial distribution depends on sample size for small N • The relationship between the distribution of sample proportions and the distribution of sample counts

Demo 15: Two-Dimensional Random Walks
Unexpected behavior in 2D random walks • How the 2D walk eventually looks like a 2D normal distribution

Standard Deviation, Standard Error, and Student’s t

Demo 16: Standard Error and Standard Deviation
Getting a feel for the difference between standard deviation and standard error

Demo 17: What Is Standard Error, Really?
The connection between standard error and the sampling distribution of the mean • How the sample size connects standard deviation and standard error

Demo 18: The Road to Student’s t
Using standard error as the scale for measuring how far a sample mean is from the true mean • How these quantities are not normally distributed; in fact they follow a t-distribution

Demo 19: A Close Look at the t-Statistic
How sample mean, standard deviation, t, and P interrelate • How they depend on the values of individual points in a sample

Sampling Distributions

Demo 20: The Distribution of Sample Proportions
How sample size and population proportion affect the distribution
Contents

Demo 21: Adding Uniform Random Variables 61
What happens when you add two uniform random variables • How that corresponds to adding two dice

Demo 22: How Errors Add 63
Basic error analysis • How to find the error in the sum of two quantities that each have some measurement error

Demo 23: Sampling Distributions and Sample Size 65
How sampling distributions (of the mean) get narrower as you increase sample size

Demo 24: How the Width of the Sampling Distribution Depends on \( N \) 68
How the width (as measured by IQR) of a sampling distribution of the mean is inversely proportional to the square root of the sample size

Demo 25: Does \( n – 1 \) Really Work in the SD? 71
Unbiased estimators • How the familiar formula for sample standard deviation is not unbiased • Why we should care about variance

Demo 26: German Tanks 73
Unbiased estimators • Evaluating estimators from their sampling distributions • Even among unbiased estimators, some are better than others

Demo 27: The Central Limit Theorem 76
A demo of the CLT • How sampling distributions usually look normal • Cases where they do not look normal

Confidence Intervals 79

Demo 28: The Confidence Interval of a Proportion 80
Defining the confidence interval • Looking at sample results in terms of plausibility

Demo 29: Capturing with Confidence Intervals 83
How confidence intervals of a proportion do not always capture the population value

Demo 30: Where Does That Root (\( p(1 – p) \)) Come From? 85
The standard deviation of a variable that’s only 0 or 1 • Connecting the “proportion” situation to the “mean” situation

Demo 31: Why np > 10 Is a Good Rule of Thumb 87
Explaining the np > 10 rule for using the normal approximation in the CI of a proportion

Demo 32: How the Width of the CI Depends on \( N \) 90
How the width of a confidence interval is inversely proportional to the square root of the sample size

Demo 33: Using the Bootstrap to Estimate a Parameter 93
The bootstrap • Using resampling (with replacement) to create an interval for a parameter
Demo 34: Exploring the Confidence Interval of the Mean
How the CI depends on individual values

Demo 35: Capturing the Mean with Confidence Intervals
How confidence intervals of a mean do not always capture the population value
What repeated CIs look like

Hypothesis Tests

Demo 36: Fair and Unfair Dice
Creating a measure of “fairness” • Sampling distributions • Testing hypotheses empirically • The chi-square statistic

Demo 37: Scrambling to Compare Means
Randomization test • Using scrambling to simulate the null hypothesis • Generating a sampling distribution

Demo 38: Using a t-Test to Compare Means
Comparing means with Student’s t

Demo 39: Another Look at a t-Test
Repeated t-tests on samples from the same distribution • How t, P, mean, and standard deviation interrelate

Demo 40: On the Equivalence of Tests and Estimates
How a hypothesis test and a confidence interval are really the same

Demo 41: Paired Versus Unpaired
How a paired test gives a significant result more easily than its unpaired counterpart

Demo 42: Analysis of Variance
Assessing whether means are different in different groups • Introduction to ANOVA

Power in Tests

Demo 43: The Distribution of P-Values
How the distribution of P is flat if the null hypothesis is true • How it changes if the null hypothesis is false

Demo 44: Power
How power—the chance that you reject the null hypothesis—changes with the population parameters

Demo 45: Power and Sample Size
How power—the chance that you reject the null hypothesis—changes with sample size