Fifty Fathoms • Selected Responses

In "The Meaning of Mean":

2 (p. 23) There are 5 points in group A and 25 in group B. How do those quantities relate to the distances from the mean?

The overall mean is five times as close to the **B** mean as the **A** mean. You could also think of it this way: the arithmetic mean of all the values is a weighted average of the two group means, with the weights being 25 and 5.

In "Mean and Median":

3 (p. 24) Why does the median stick to the moving point when you move it past the median?

Over that range, the median *is* that value. Later, we'll add a sixth point—so there will be an even number of data points. Then this will no longer be the case.

11 (p. 25) How far do you have to move the right-hand point to increase the midrange by one?

Two units, as long as it's the right-hand point. The midrange is the average of the minimum and the maximum, so to get their average to increase by one, you need to increase one of them by two.

14 (p. 25) How would you characterize how "resistant" the midrange is to changes in outliers—especially compared with the mean and median?

The mean changes when you drag any point. The median changes only when you drag the point in the middle. The midrange changes only when you drag a point on the outside. So in a way, it's the least resistant to changes in outliers.

In "What Do Normal Data Look Like?":

7 (p. 27) In the normal quantile plots, (choose Normal Quantile Plot from the pop-up menu in the graph) where do the points most deviate from the straight line?

At the ends. This corresponds to the tails of the distribution; out there, where there are fewer points, random variation is more apparent in the plot.

In "Transforming the Mean and Standard Deviation":

6 (p. 29) Do the same for the middle point in the distribution. How far do you have to move it to increase the mean by one? How far to increase the standard deviation by one? Compare that to moving the extreme point and explain why it's the same or different.

Standard deviations behave differently from means. A change in an extreme value affects the SD more than a change in a central one, because a small change in a large deviation makes a bigger difference in the square. For example, a change from 1 to 2 increases the square by 3, but a change 10 to 11 increases the square by 21.

Of course, a change in one value changes the mean, so dragging one point changes all the other deviations as well. But *these* changes will be the same (per unit drag) whether the moving point is in the center or in the tail.

In "The Mean is Least Squares, Too":

1 (p. 32) What special value minimizes the sum of absolute differences?

It turns out to be the median, of all things.

By the way, we had an odd number of values; if there are an even number, the region between them is flat—there is no unique minimum. That business of "take the average of the middle two" is just a convention.

4 (p. 32) Suppose that instead of adding up squares or absolute differences, we minimized the (absolute) difference between the candidate and the farthest data point. What would the graph look like? What value would you get for a minimum?

Here you get the midrange—the place where the distance from the point to the two the extremes is the same. Any closer to one, and the other becomes the farthest data point. The graph is a V, like an absolute-value function.

In "Least-Squares Linear Regression":

3 (p. 35) Why do you suppose they used squares for this?

One important reason is that if you just take the residuals (the vertical differences between the points and the line), they can be positive or negative. When you add them, a big positive residual might be offset by a large negative one. If you square them, they'll all be positive.

Another reason is to emphasize the contribution of the points far from the line; these points "pull harder" on the line than the close ones. If that's not what you want, you might rather use a different line such as the median-median.

In "Devising the Correlation Coefficient":

1 (p. 39) When you move a point, the others (on the zx-zy graph) first move in the opposite direction. Why?

If you make a value higher, you're raising the mean. And as the mean rises, the *z*-scores of the other points fall.

2 (p. 39) As you move the point even farther in one direction, what happens to the other points?

They collapse towards one another in the direction of motion. That is, if you're moving a point a long way horizontally, the other points will form a vertical line. This is because the moving point makes the SD so large. Since SD is in the denominator, all the other cases get a very small *z*-score.

9 (p. 39) How did the correlation coefficient—the mean of the prods—change as you dragged the points to the center? Why?

It got smaller. When one point is extreme and the others are clumped, the correlation is near +1 or -1. When you finally bring the extreme point in, the SDs suddenly get a lot smaller—which expands the *zx-zy* graph. Now we can see the irregularities in the clumping; the points are more randomly scattered, and show a small correlation.

In "Correlation Coefficients of Samples":

5 (p. 43) It looks as if the spreads for pop_corr = +0.8 are smaller than the ones for pop_corr = –0.4. Why would that be?

As the population correlation gets more extreme, the spread among the samples decreases. Here's one way to think about it: suppose the population correlation were 1.0. Then the samples would be perfectly lined up, and all have 1.0 correlations. This changes *gradually* between $r = 1$ and $r = 0$, where you can get a wide variety of correlations.

In "Regression Towards the Mean":

2 (p. 45) When you flip coins and get six heads in a row, the chance of tails is still one half if it's a fair coin. Yet here, it looks as if when you get a good score you are more likely to get a lower score next time. What is it about this situation that makes it different?

It really isn't different. Think of heads as 1 and tails as 0. You've gotten a bunch of heads (1.0) in a row. What do you expect for the next flip? The expected value is 0.5—fifty-fifty heads and tails. So you *do* expect a decrease; the gambler's fallacy is to expect *zero*.

In "Flipping Coins—the Law of Large Numbers":

3 (p. 49) Why is graph so much flatter at the end than it is at the beginning?

We're calculating proportions. From step to step, we're changing the numerator by one; if the denominator is several hundred, that doesn't make as big a change as at the beginning when the denominator is very small.

In "How Random Walks Go as Root N":

2 (p. 53) How does the average absolute value at the end of the walk (endAbs) change? (as the number of steps increases)

The mean of **endAbs** increases. As we will see, it increases less than linearly. In fact, it goes as the square root of the number of steps. It's different from the *mean* of **end** because the values for **end** include negative numbers that roughly balance the positives.

In "How Random Walks Go as Root N":

6 (p. 55) If the mean square distance is N, does that mean that the mean absolute distance is the square root of N? Exactly? Not at all? Only in the limit? Only on Tuesdays? (One way to look at it: in the case table, create a new attribute equal to the square root of N; then replace N with this new attribute in the first (N-endAbs) graph. The line is straight; what is its slope? How do you explain that it is not 1.0?)

Not exactly; it looks as if the slope is about 0.8. Now, these distances are roughly normally distributed, especially for large *n*. And, in fact, the root-mean-square distance is about 5/4 the mean absolute distance for normal data.

To convince yourself that this is even possible, consider a data set with two values, 0 and 2. The sum of the squares is 4, so the mean square distance is 2; the root-mean-square distance is $\sqrt{2}$, or about 1.4. But the mean *absolute* distance is 1.

In "Building the Binomial Distribution":

3 (p. 58) Why should the spread be smaller when p_pop is close to 1 or 0?

Here is one way to think of it: we have twenty cards. If $p = 0.5$, there will be roughly 10 blacks, but it would not be hard to get 12 (a difference of 2, or 0.1 in the proportion). But suppose $p = 0.95$. We expect 19 blacks. But to get 17 (also a difference of two) you have to hit two more of those rare reds.

In "More Binomial":

4 (p. 60) How can you predict where the peak will be in the howMany graph? How about the p_hat graph?

The **howMany** peak will be at about **p * N**. For **p_hat**, the peak will be at about **p**. If **N** is small, there may be no proportion near **p**, so the peak will be on the nearest "stack," or may straddle two bins.

In "Two-Dimensional Random Walks":

1 (p. 62) It should be pretty clear that, since each step is only one unit long, all points must lie in a disk defined by $r \le 2$, where r is the distance to the origin. What would the two graphs look like if the distribution of points were uniform in that disk?

The left-hand graph would show a roughly uniform scatter of points. But the right-hand graph (the histogram by radius) would appear to increase *linearly* with radius, passing through the origin. That's because the height of the bar—the population—is proportion to area, and the area of the ring represented by each bar is proportional to its circumference, and $C = 2\pi r$.

2 (p. 63) Use that result (or anything else you can think of) to explain why the obvious clump near the origin doesn't show up in the radial histogram, and why the "density ring" at r = 2 is so striking in the histogram but not in the scatter plot.

The central clump manifests itself in the histogram by being higher than it would be if it were straight linear. That is, the plot really *does* show up in the histogram. As to the upswing at $r = 2$, maybe it doesn't look as striking on the scatter plot because it's on the edge of the points—but it's clearly there in the histogram.

In "What Is Standard Error, Really?":

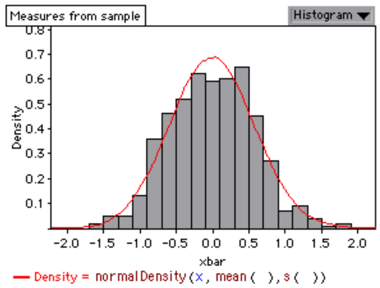
2 (p. 70) What standard deviation do you expect in the sampling distribution if there are 16 cases in each sample, and the original standard deviation is 1.0? Try it and see if you're right.

We should get $\sigma/\sqrt{N}$, which is 1/4.

3 (p. 70) With the sampling distribution as a histogram, make Fathom draw the relevant normal curve so you can compare. (You may want to use a density scale on the graph—in the Graph menu, look under Scale; you can what it looks like in the next demo.)

With that density scale, use the formula
NormalDensity(x, mean( ), s( )),
as shown in the illustration. We're plotting sample means; the **mean( )** is the mean of those means; the **s( )** is their sample standard deviation.

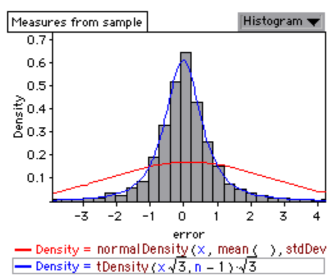
This one is from $n = 3$:



In "The Road to Student's t":

3 (p. 74) Why didn't we just measure the difference from the true mean in standard deviations instead of standard errors? It would be dimensionless—independent of the original units—and in a scale that depended only on the sample we got. Try it, see what happens, and explain your results.

Again, the large amount of data in the tails of this distribution makes the normal density function with the same SD much broader. A little serious reflection tells us that we again have a t distribution, but scaled differently because of the sample size. An example with $n = 3$:



In case you can't read it, that density function is

$$tDensity(x\sqrt{3},\ n-1)\sqrt{3}\ .$$

That is, a test statistic of 2 sample standard deviations is as unusual as a statistic of $2\sqrt{n}$ SEs (i.e., $t = 3.46$ when $n = 3$); it's in the same place in the distribution.

In "A Close Look at the t Statistic":

2 (p. 76) When you moved the point to the right, why did t decrease? Isn't the mean getting farther from zero?

The mean is getting farther from zero, and that tends to make t increase. But the SE is also getting bigger—which makes t smaller. If the point's data value is large enough, this overcomes the effect of the increased mean.

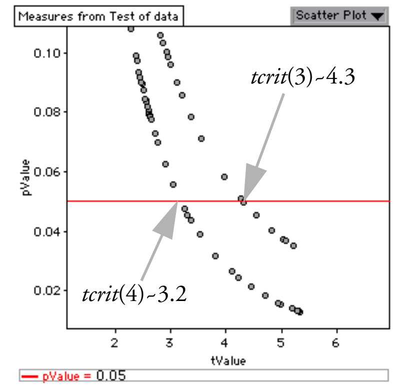
Remember that the mean moves at $1/n$ the speed of a single data point ("The Meaning of Mean" on page 22). The SE increases at a faster rate, if the moving point is far enough out.

4 (p. 76) Which point do you have to move to get the smallest standard deviation? Why? (Trick question alert.)

Move either point 0 or point 2. To get the smallest SD, move it between the other two. Moving point 1 doesn't cut it: because it's in the middle, you can't get rid of the spread between 0 and 2, giving you a larger SD.

3 (p. 77) As in the first task, zoom in to see where the critical values for t are. There are now two—one for a sample of three, one for four. Confirm that the one for four is smaller than the one for three and explain why.

Here is such a graph, all zoomed in:

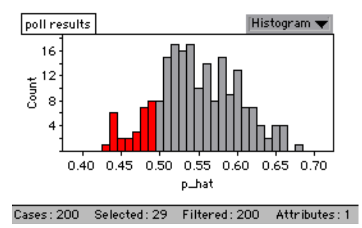


The diagram shows how the critical value for 4 is smaller than the one for 3. But why? One way to look at it is to remember the shape of the t distribution itself. It has long tails compared to the normal, and the larger the sample size, the more normal it is—and the smaller the tails. We want to know where the 97.5 percentile is (this is a two-tailed test), so that will be *larger* when $n = 3$ (more weight farther out in the tails) than when $n = 4$.

In "The Distribution of Sample Proportions":

4 (p. 80) If you poll a sample of 100 people from a town, and in reality, 55% of the town wants new sewers, it is possible that less than half your sample will say they want new sewers. About how likely is that?

In Fathom, set **N**=100 and **p**=0.55. Then you have several strategies. One is to select all of the cases below 0.5, point at the collection, and look in the status bar to see how many cases are in the tail below 0.5:



Here we see that 29/200, or 14.5% of the time, our poll will suggest that less than half the town wants new sewer, when in fact, 55% do. The precise numbers will vary. (Theoretically, we expect about 13.5%.)

In "How Errors Add":

1 (p. 84) It looks as if the function is pretty straight if you get far from the y-axis. Find the equation for that line and explain why it has to be that way—using reasoning about errors.

You can also look at the question as being, how does

$$y = \sqrt{1 + x^2}$$

behave as x grows without bound? The answer is, it approaches $y = x$. This makes sense conceptually; if you add two quantities, and the error for one is huge compared to the other's error, the error in the sum will just be the big error: the smaller error is negligible.

2 (p. 84) The function looks flat as you get close to the y-axis. Explain that too.

It's the opposite situation. If you have a tiny error compared to your partner, you can add a little and make virtually no difference in the error of the sum. But as soon as one error is not negligible compared to the other, they both make appreciable contributions.

In "Sampling Distributions and Sample Size":

2 (p. 86) What do you notice about the distributions of the sample means when you increase the sample size?

The distribution gets tighter (narrower) as sample size goes up.

(p. 87) Do the same as in the previous extension, but for standard deviation (use sampSD) instead of median. In this case, the distribution gets narrower, as before, but something else happens as well. What is it?

The center of **sampSD** is systematically low for **N = 5**. This is because the sample standard deviation—even using $n - 1$ in the denominator—underestimates the population standard deviation.

In "How the Width of the Sampling Distribution Depends on N":

3 (p. 90) Suppose you didn't think of using the square root in the denominator—you just saw that the inverse function didn't work. What could you do, in exploring the data, to "discover" the square root relationship?

One strategy is to use logarithms. Create new attributes that calculate the log of **N** and the log of **S2**. When you plot them against one another, the result will be linear. In this case, the slope should be about –0.5, which would indicate that **S2** is inversely proportional to the square root of **N**.

You can see how to use logs in an investigation in "How the Width of the CI Depends on N" on page 110 (it's in an extension near the end).

In "Does n – 1 Really Work in the SD?":

2 (p. 92) How could we have predicted that the mean of the "plain" SD distribution would be smaller than the first one we tried?

The plain SD has n in the denominator instead of $n - 1$. So each sample SD is slightly smaller.

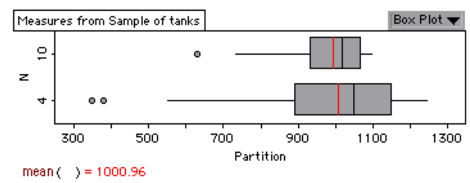
In "German Tanks":

2 (p. 94) What are the largest and smallest possible values for twiceMean?

We're sampling with replacement, so the smallest average is 1 and the largest is 1000—which give estimates of 2–2000. If we were sampling *without* replacement, the smallest sample would be {1, 2, 3, 4}. That mean is 2.5, so the estimate would be 5. The largest is {997, 998, 999, 1000}, which gives an estimate of 1997.

6 (p. 95) Rerun the simulation with a sample of ten tanks instead of only four. How do the distributions of the estimators change? To change that sample size—to capture more tanks—open the inspector for the Sample of tanks (opened) collection, and change the number sampled from 4 to 10.

The means of the estimates are about the same, but their distributions—especially **Partition**'s—get tighter. Here are box plots of the simulation run with sample sizes of 4 and 10:

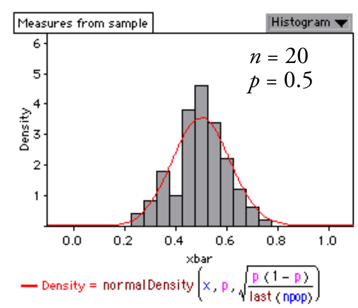


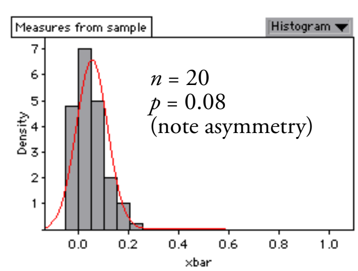
In "The Central Limit Theorem":

2 (p. 97) Imagine making an asymmetrical, sharply bimodal distribution (just zeros and ones, but lots more zeros). Think about what we found out about what it takes to make the sampling distribution of the mean look normal. How does that relate to the rule of thumb (np > 10) for using the normal approximation when you calculate confidence intervals for proportions? (See "Why np>10 is a Good Rule of Thumb" on page 107.) In fact, don't just imagine it; do it!

Both situations are about sampling distributions. If the probability of getting a one is p and you choose n for your sample size, you'll get a binomial distribution with those two parameters. You can approximate the distribution of the sample mean using a normal distribution with mean p and SD $\sqrt{p(1-p)/N}$. If np is small, however, this normal approximation will not be very good. The distribution will often be asymmetrical—and obviously not normal.

Here are two sampling distributions, labeled to show the relevant values:





This does not address the "choppiness" of normal quantile plots when n is small. Even at $n = 40$, it's easy to see that the distribution of **xbar** is discrete, not continuous.

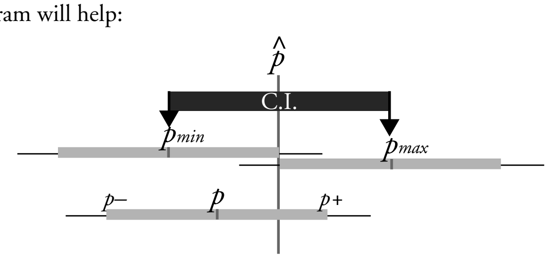
In "The Confidence Interval of a Proportion":

1 (p. 102) Explain clearly why this "plausibility interval" is really the same as the orthodox confidence interval ("if you were to draw new samples and construct intervals repeatedly…") described at the beginning of this demo.

Suppose the true population proportion is p. Now, 95% of the time, the sample proportions will be within some interval—from p_- to p_+. If we show that our process always captures the true proportion whenever the sample proportion is in that interval (and not otherwise) that means that the process does so 95% of the time—which is the orthodox requirement.

So now: whatever $\hat{p}$ we get, we find p_{max}, the p such that $\hat{p}$ is at the *bottom* of its 95% band; and p_{min}, where $\hat{p}$ is at the *top*. (That's our process.) OK: suppose $\hat{p}$ is in the interval $[p, p_+]$, as shown. Then p_{min} must be less than p, because if it were greater than p, that would mean that $\hat{p}$ had to be greater than p_+ (because of the way p_+ was constructed). Therefore we captured the population proportion. We use this same reasoning for $\hat{p} < p$, and the cases outside the interval.

This seems almost doubletalk, but it's really not. If you're confused (and this author wouldn't blame you) maybe this diagram will help:



In any case, the plausibility interval—the range of population parameters over which the test statistic is plausible—does in fact exactly satisfy the requirements of the C.I.

In "Where Does That Root(p(1 – p)) Come From?":

3 (p. 106) Does that mean that the confidence interval there has zero width? Why or why not?

No. It may still be plausible for the population to have some probability for success ($p > 0$), but that this particular sample had no successes in it ($\hat{p} = 0$). This is an illustration of the fact that the normal approximation does not apply here.

5 (page 106) Show that the variance is really $p(1 - p)$.

We'll call our random variable X, which is 0 with probability $(1 - p)$ and 1 with probability p. The mean of X—its expected value—is p. So the variance is the expected value of $(X - p)^2$. We find that by summing over both possible values, weighting by probability:

$$\begin{aligned} Var[X] &= E[(X-p)^2] \\ &= P(X=0)(0-p)^2 + P(X=1)(1-p)^2 \\ &= (1-p)p^2 + p(1-p)^2 \\ &= (1-p)[p^2 + p(1-p)] \\ &= (1-p)p. \end{aligned}$$

The SD is the square root of that.

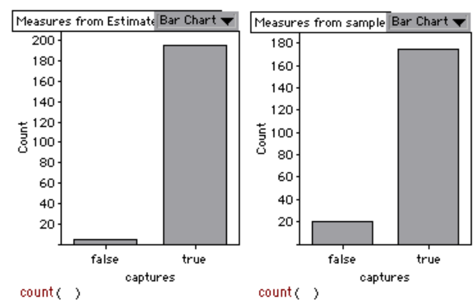
In "Why np>10 is a Good Rule of Thumb":

3 (p. 109) If you take the sample size and multiply it by the sample proportion, don't you get the number of successes? That is, could you say that the rule of thumb "np>10" really means that you have to get at least 10 people to say yes (and at least 10 people to say no) before you can use the normal approximation to the CI?

Yes! That also means that to get np>10 and $n(1 - p)$>10, you must have a sample size of more than 20.

6 (p. 109) Perform an experiment to test how much difference it makes to use the normal approximation instead of the binomial. Use a sample size of 10 and a true population proportion of 0.2. Draw repeated samples and construct 95% confidence intervals using both the binomial estimate and the normal approximation. See how many CIs capture the true proportion, and on which side they miss. Vary the sample size and the population proportion.

We see that Fathom's CI (which uses a binomial distribution) is conservative—it captures more than 95%. The normal approximation, however, captures only about 90%, as shown in these two graphs of 200 simulations each:



The left-hand graph shows Fathom's binomial-based estimate; the right-hand is from the normal approximation.

But why is using the binomial "conservative?" Doesn't it give the exact answer? The binomial distribution is the correct distribution, but it is discrete, so it may not be possible to encompass *exactly* 95%. So we err on the side of conservatism and include enough values to encompass *at least* 95%.

In "How the Width of the CI Depends on N":

1 (p. 111) How large a sample would we need to reduce the "margin of error" to 0.01? That is, how many points do we need in order to get confidenceWidth as low as 0.01?

The **confidenceWidth** function we derived in the graph—let's call it w—is roughly

$$w = \frac{2}{\sqrt{N}}, \text{ or}$$

$$N = \frac{4}{w^2}$$

So to get $w = 0.01$, we need 40,000 points.

6 (p. 111) What does it mean that the values of confidenceWidth get closer together as sampleCount increases?

It does *not* signify that we know the mean better! That's because the *value* of those **confidenceWidths** is small.

Rather, this shows how the variation among confidence widths decreases as sample size increases. If we cared about something we might call "the confidence interval of the width of the confidence interval," this would be it.

In "Using the Bootstrap to Estimate a Parameter":

1 (p. 114) If the median of the original sample is 7453, how can it be that so many of the bootstrap samples have a median of 8000? Shouldn't 7453 be, if not the most popular value, at least close?

It seems that 8000 is a common number in the original sample. If we take bootstrap samples, the number 8000 often shows up as the median; it depends which cases get duplicated or left out by the resampling process. Because 8000 is common *in the vicinity of the true median*, it will appear more frequently as the bootstrap median than it does in the original sample. 7453 happens to be the original median, but it doesn't even appear in the original sample (there are an even number of cases)—so it appears only seldom as the bootstrap median.

3 (p. 114) Explain why the median income of Champaign County seems to be so low. Can it be that half the people earned less than $7500 per year, even way back in 1989?

Yes; the data include a huge number of zeroes—in fact, these data include children. (The **age** attribute is in the data set, so you can remove them with a filter if you wish.)

In "Exploring the Confidence Interval of the Mean":

4 (p. 116) Where are the points when the interval does not include zero?

All on one side, and close together.

7 (p. 117) Getting a point outside the CI is hard with three points. Figure out what has to be true for that to occur.

First, you have to make a point extreme. Interestingly, you can do that by making the points {0, 0, 1}. This works because the "1" is extreme *in relation to the spread of the points*—and it's just as extreme as {0, 0, 1000}. But it's still not outside the interval. So shrink the interval by lowering **CI_pct**.

In "Capturing the Mean with Confidence Intervals":

2 (p. 119) In general, what characterizes the intervals that miss (besides the fact that they miss)? That is, what do most of them have in common?

They're *short*.

4 (p. 119) As you drag CI_pct to lower values, what happens to the number of intervals that miss? Explain why that is, based on watching the display of the 50 intervals?

As **CI_pct** goes down, the intervals get shorter (but stay in place); as this happens, more of them cannot reach the true value, and miss.

This makes sense; we expect an 80% CI to miss the true value more than a 95% CI.

5 (p. 119) When we looked at intervals of proportions in "Capturing with Confidence Intervals" on page 103, the intervals were all roughly the same length. Here they differed. Why? Here, the short ones tended to miss. There, we got the same proportion of misses even though none were so short. How could that be?

With this small sample, the SD of the samples varies widely—so we get long and short CIs. With proportions, the SDs are all about the same— $\sqrt{p(1-p)}$ —so the CIs are about the same width. But the same proportion miss regardless; that's how they're constructed.

In "Fair and Unfair Dice":

2 (p. 124) What is the largest possible value you could get for evenMinusOdd? Explain why.

108. Suppose you rolled only even numbers…

5 (p. 125) Does the chiSquare statistic seem more unusual, less unusual, or about the same as the evenMinusOdd statistic?

ChiSquare is less unusual, or less significant—there are more points in the tail beyond the test statistic—than the **evenMinusOdd** statistic. This is probably because **evenMinusOdd** focus on exactly why the test die is strange, whereas **chiSquare** will pick up a wider variety of nonuniformity.

In "Scrambling to Compare Means":

3 (p. 127) Is our test statistic of +1.0 particularly unusual, or is it the sort of thing that arises by chance?

The +1.0 comes up once in awhile; and if you include the –1.0s and below, it's pretty common.

5 (p. 128) What result would you get if the original data had disjoint groups—that is, the highest value for one group was lower than the lowest value for the other?

The test statistic would probably be at the extreme of the distribution.

Disjoint data gives you the maximum possible difference of means for your test statistic. So the only way that value could even *appear* in the sampling distribution would be if the data got shuffled into their original groups: possible, but unlikely with this many data points.

In "Using a t-Test to Compare Means":

1 (p. 129) How can you predict, before you drag a given point, which way the P-value will change?

If you're dragging it in a direction that makes the groups more different, the *P*-value will go down. At least at first. But see this next answer…

3 (p. 130) Explain how that is possible. That is, by dragging a point from the higher group up, you're making the two groups more different, aren't you? So it makes sense that p should decrease. But then why does it increase again?

As you drag the point up, how does it affect *t*? It increases the mean of the group, which increases *t*. But eventually, it also increases the SD of the group. As the point becomes more and more an outlier, that increase in the SD overwhelms the increase in the mean—so it *decreases t*.

In "Another Look at a t-Test":

1 (p. 132) In the first graph, why are there points only in the first and third quadrants? Is it possible to have points anywhere in those quadrants?

The sample SDs can be very different—that changes *t*, and therefore *P.* This also answers the previous question: the points can be *anywhere* in the first or third quadrants—just some situations (small mean, large *t*) are unlikely.

2 (p. 132) How can it be that one sample with a mean of 0.6 can have a smaller (i.e., less significant) value for t than a different sample with a mean of 0.4?

5 (p. 132) Can you tell from a point on the graph whether it was made with mu = 0 or not?

No.

In "On the Equivalence of Tests and Estimates":

4 (p. 134) Why does the right-hand rectangle go nuts (e.g., turn into a fat gold ball) sometimes?

5 (p. 134) Why doesn't the left-hand rectangle go nuts at the same time?

If you set **alpha** to less than zero, Fathom sets the confidence level to a number greater than 100%. Then Fathom cannot compute a confidence interval (you're asking Fathom to come up with an interval that will enclose the true mean more than always). But note that changing alpha has no effect on the hypothesis test. That's because it produces the same *P*-value regardless of the alpha level; *we* do the comparison to tell if it's significant. If the alpha level is negative, the rectangle will always be black—no significant difference—because the (always-positive) *P*-value is greater.

In "Paired Versus Unpaired":

3 (p. 136) How about in the unpaired test?

You can't see the mean improvement directly, but you can see the means of the two groups. Subtract them to get the difference of means (which is the same as the mean improvement in the paired test).

7 (p. 136) Suppose you're designing a study to see whether students' attitudes towards drug use change after they have seen a series of films. Your main measurement instrument is a questionnaire. What are the main arguments for and against having students put their names on it?

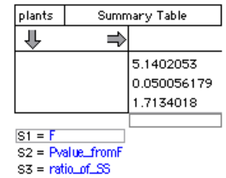
If you have names, you can pair the questionnaires, and do a paired test—which is more powerful. But drug use is a tricky subject; if students put their names on the questionnaires, they might not be candid (i.e., they might lie), and your "data" will be worthless.

In "Analysis of Variance":

2 (p. 138) In the lower summary table, the right-hand column has a bunch of zeros in it. Why are those numbers zero?

The zeros are the sums of the squares of differences of the **gHeight**s in the group from the mean of the **gHeight**s. But **gHeight** *is* that mean, so we're just adding zeros.

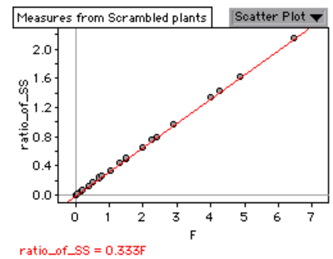
4 (p. 140) What value for F gives a P-value of about 0.05? (i.e., What is the critical value for F at the 0.05 level?)



Using the summary table, and dragging points, you can find *P* of about 0.05 with an *F* of about 5.14, as shown in this illustration:

This should match the critical value you'd find in a table, and it does. (Look it up with the two different *dfs*: 2 and 6.)

8 (p. 140) Plot F against ratio_of_SS. Explain the graph you see.



The graph has a slope of 1/3, which shows that, indeed, *F* is proportional to the ratio we constructed. The "3" is the ratio of degrees of freedom: 6 : 2.

In "Power":

2 (p. 146) What would happen to these curves if the population standard deviation were 2.0 instead of 1.0?

In that case, it would take twice the difference in **trueMu** to create the same effect: the curves would rise twice as far from the vertical axis.

3 (p. 146) The points in the graph look as if they don't quite lie on a smooth curve. Why not?

They are derived from a sampling procedure, so there is sampling error.

In "Power and Sample Size":

2 (p. 148) What would happen to these curves if the population standard deviation were 2.0 instead of 1.0?

The graph would lower: all other things being equal, you get less power the larger the population SD. Another way to look at it: it's like having **trueMu** only half what it was before—after all, everything should scale with the spread of the population—so it's that much harder to detect.

Note that when we study *proportions* (e.g., polling), the SD doesn't change much—it only depends on *p*.

In "Heteroscedasticity and its Consequences":

1 (p. 151) Explain qualitatively why the slopes in the "butterfly" distribution would give you too many significant results, and why the slopes in the "diamond" distribution would give you too few.

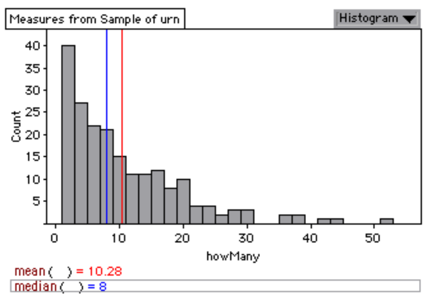
One way to think of it: least-squares lines try to go through the means of the values in vertical strips. Picture three strips: one in the middle, and one on each end.

In the butterfly, the mean in our middle strip is pretty certain, but the ones on the ends will vary more. In contrast, in the diamond, the middle strip varies, but the ones on the ends are tied down. In this way we can see that the butterfly will get more lines with a larger absolute slope. The homoscedastic, independent case is in between the two; since *P*-values are based on independence, the butterflies will naturally have a larger number of significant slopes, and diamonds will have fewer.

In "Wait Time and the Geometric Distribution":

6 (p. 156) Is the median of a geometric distribution larger, smaller, or the same as the mean? Use Fathom to demonstrate that what you say is correct. Also, explain why it is correct based on the shape of the distribution.

The median is smaller, as is usually the case for distributions with a tail that extends to the right. If it were symmetrical, the mean and median would be the same. Move some of the right-hand points farther to the right, and you increase the mean without increasing the median. In Fathom, just plot the **median( )** as well as **mean( )** to see:



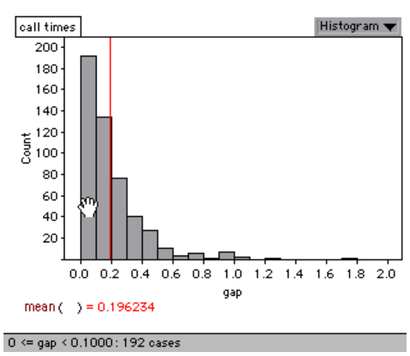
7 (p. 156) What is the mode of a geometric distribution? Explain.

One! You are most likely to succeed—draw the red ball, roll a six, whatever—on your first try in this wait-time situation. This is counterintuitive, since we just proved that the *expected* value is larger. But it is correct; all it is saying is that you are more likely to succeed on the first try than on any other *single, specific* turn.

In "The Exponential and Poisson Distributions":

2 (p. 159) In that situation, after what fraction of the calls (out of 500 calls) was there a gap—the time between calls—of six seconds or less? (Six seconds is 0.1 minutes.)

This shows the distribution with **averageTime**=0.2. With the hand over the left-hand bar of the histogram (which is the range from zero to six seconds), we can read the number in the status bar: 192. Your results will be different due to random variation.



So that's 192/500, or about 38% of the calls.

3 (p. 159) If we average one call per minute (so we'll be simulating about 500 minutes), in about how many minutes do you expect to get zero calls? Remember, you'll have to think theoretically—or be sneaky—as Fathom will not display the data.

Let's be sneaky. The illustration shows the bottom of one of our case tables—the one derived from the summary table:

Measures from call times Table

	minute	S1
317	500	1
318	501	1
319	503	1
320	504	1
321	506	1
322	507	1
323	508	1

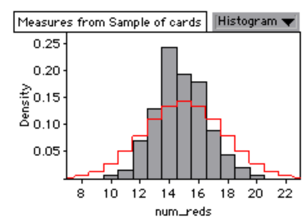
It shows, for example, that there was one call during minute number 508—the last minute on the table. You can also see, for example, that no calls came in during 505.

Thinking about this table for a moment, we realize that there are 323 rows representing 508 minutes. That means that (508 – 323) minutes are *not represented*: 185 minutes had no calls.

In "Sampling Without Replacement and the Hypergeometric Distribution":

2 (p. 161) At n = 30, how would you characterize the difference between the data from the samples and the theoretical curve?

The data are more tightly clustered, closer to the expected value of 15 reds. The line in the illustration shows the binomial distribution for comparison.



4 (p. 161) If we sampled 52 cards instead of 30, what would the distribution look like?

The distribution would be a spike at 26.

5 (p. 162) Make a Fathom document to simulate card counting in this scenario: (etc.)

The "obvious strategy" we used is: if the majority of cards played so far are red, bet black, and vice versa.

See **Draw 21 cards.ftm** in the **other files** folder. This graph showing the results of 50 trials of 100 games. You *can* lose, but on the average, you win about $10 in every 100 games played.