R version 3.3.0 (2016-05-03) -- "Supposedly Educational"

Copyright (C) 2016 The R Foundation for Statistical Computing

Platform: x86\_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.

You are welcome to redistribute it under certain conditions.

Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.

Type 'contributors()' for more information and

'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or

'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

[Previously saved workspace restored]

> library(Swirl)

Error in library(Swirl) : there is no package called ‘Swirl’

> library(swirl)

| Hi! I see that you have some variables saved in your workspace. To keep things running

| smoothly, I recommend you clean up before starting swirl.

| Type ls() to see a list of the variables in your workspace. Then, type rm(list=ls()) to

| clear your workspace.

| Type swirl() when you are ready to begin.

Warning message:

package ‘swirl’ was built under R version 3.3.1

> swirl()

| Welcome to swirl! Please sign in. If you've been here before, use the same name as you

| did then. If you are new, call yourself something unique.

What shall I call you? SY

| Please choose a course, or type 0 to exit swirl.

1: Statistical Inference

2: Take me to the swirl course repository!

Selection: 1

| Please choose a lesson, or type 0 to return to course menu.

1: Introduction 2: Probability1 3: Probability2

4: ConditionalProbability 5: Expectations 6: Variance

7: CommonDistros 8: Asymptotics 9: T Confidence Intervals

10: Hypothesis Testing 11: P Values 12: Power

13: Multiple Testing 14: Resampling

Selection: 6

| Attempting to load lesson dependencies...

| Package ‘ggplot2’ loaded correctly!

| Package ‘jpeg’ loaded correctly!

| | | 0%

| Variance. (Slides for this and other Data Science courses may be found at github

| https://github.com/DataScienceSpecialization/courses/. If you care to use them, they must

| be downloaded as a zip file and viewed locally. This lesson corresponds to

| 06\_Statistical\_Inference/05\_Variance.)

...

| |== | 2%

| In this lesson, we'll discuss variances of distributions which, like means, are useful in

| characterizing them. While the mean characterizes the center of a distribution, the

| variance and its square root, the standard deviation, characterize the distribution's

| spread around the mean. As the sample mean estimates the population mean, so the sample

| variance estimates the population variance.

...

| |=== | 4%

| The variance of a random variable, as a measure of spread or dispersion, is, like a mean,

| defined as an expected value. It is the expected squared distance of the variable from

| its mean. Squaring the distance makes it positive so values less than and greater than

| the mean are treated the same. In mathematical terms, if X comes from a population with

| mean mu, then

...

| |===== | 6%

| Var(X) = E( (X-mu)^2 ) = E( (X-E(X))^2 ) = E(X^2)-E(X)^2

...

| |====== | 8%

| So variance is the difference between two expected values. Recall that E(X), the expected

| value of a random variable from the population, is mu, the mean of that population.

...

| |======== | 10%

| Higher variance implies more spread around a mean than lower variance.

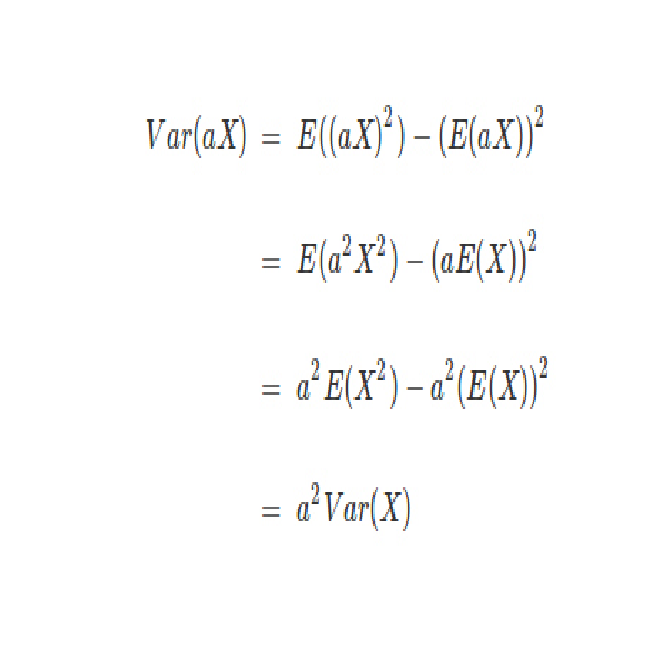
...

| |========= | 12%

| Finally, it's easy to show from the definition and the linearity of expectations that, if

| a is a constant, then Var(aX)=a^2\*Var(X). This will come in handy later.

...

 | |=========== | 13%

| If you're interested, here's the proof. You might have to stretch out your plot window to

| make it clearer.

...

| |============= | 15%

| Let's practice computing the variance of a dice roll now. First we need to compute

| E(X^2). From the definition of expected values, this means we'll take a weighted sum over

| all possible values of X^2. The weight is the probability of X occurring.

...

| |============== | 17%

| For convenience, we've defined a 6-long vector for you, dice\_sqr, which holds the squares

| of the integers 1 through 6. This will give us the X^2 values. Look at it now.

> dice\_sqr

[1] 1 4 9 16 25 36

| That's correct!

| |================ | 19%

| Now we need weights. For these we can use any of the three PDF's, (dice\_fair, dice\_high,

| and dice\_low) we defined in the previous lesson. Using R's ability to multiply vectors

| componentwise and its function 'sum' we can easily compute E(X^2) for any of these dice.

| Simply sum the product dice\_sqr \* PDF. Try this now with dice\_fair and put the result in

| a variable ex2\_fair.

> ex2\_fair <- dice\_sqr \* dice\_fair

| You're close...I can feel it! Try it again. Or, type info() for more options.

| Type 'ex2\_fair <- sum(dice\_fair \* dice\_sqr)' at the command prompt.

> ex2\_fair <- sum(dice\_fair \* dice\_sqr)

| You nailed it! Good job!

| |================= | 21%

| Recall that the expected value of a fair dice roll is 3.5. Subtract the square of that

| from ex2\_fair to compute the sample variance.

> ex2\_fair - 3.5

[1] 11.66667

| You're close...I can feel it! Try it again. Or, type info() for more options.

| Type 'ex2\_fair-3.5^2' at the command prompt.

> ex2\_fair-3.5^2

[1] 2.916667

| All that practice is paying off!

| |=================== | 23%

| Now use a similar approach to compute the sample variance of dice\_high in one step. Sum

| the appropriate product and subtract the square of the mean. Recall that edh holds the

| expected value of dice\_high.

> sum(dice\_fair \* dice\_high) - 3.5^2

[1] -12.08333

| That's not the answer I was looking for, but try again. Or, type info() for more options.

| Type 'sum(dice\_high \* dice\_sqr)-edh^2' at the command prompt.

> sum(dice\_high \* dice\_sqr)-edh^2

[1] 2.222222

| You are amazing!

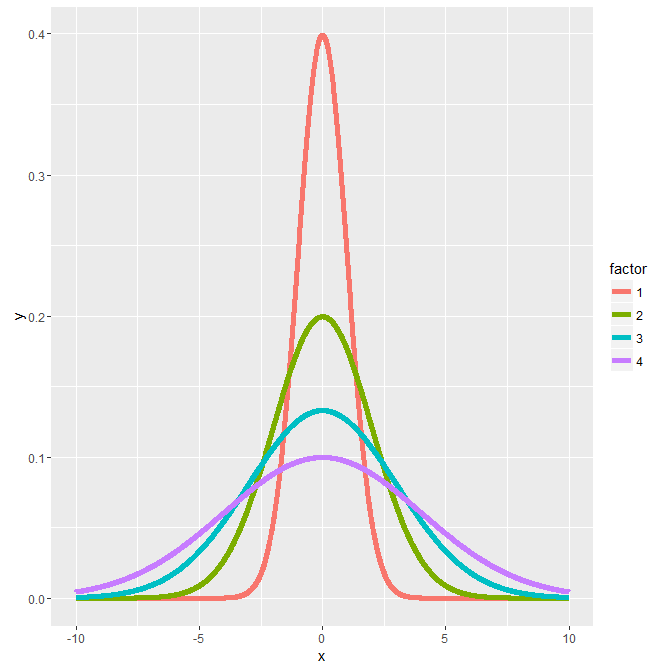
| |==================== | 25%

| Note that when we talk about variance we're using square units. Because it is often more

| useful to use measurements in the same units as X we define the standard deviation of X

| as the square root of Var(X).

...



| |====================== | 27%

| Here's a figure from the slides. It shows several normal distributions all centered

| around a common mean 0, but with different standard deviations. As you can see from the

| color key on the right, the thinner the bell the smaller the standard deviation and the

| bigger the standard deviation the fatter the bell.

...

| |======================== | 29%

| Just as we distinguished between a population mean and a sample mean we have to

| distinguish between a population variance sigma^2 and a sample variance s^2. They are

| defined similarly but with a slight difference. The sample variance is defined as the sum

| of n squared distances from the sample mean divided by (n-1), where n is the number of

| samples or observations. We divide by n-1 because this is the number of degrees of

| freedom in the system. The first n-1 samples or observations are independent given the

| mean. The last one isn't independent since it can be calculated from the sample mean used

| in the formula.

...

| |========================= | 31%

| In other words, the sample variance is ALMOST the average squared deviation from the

| sample mean.

...

| |=========================== | 33%

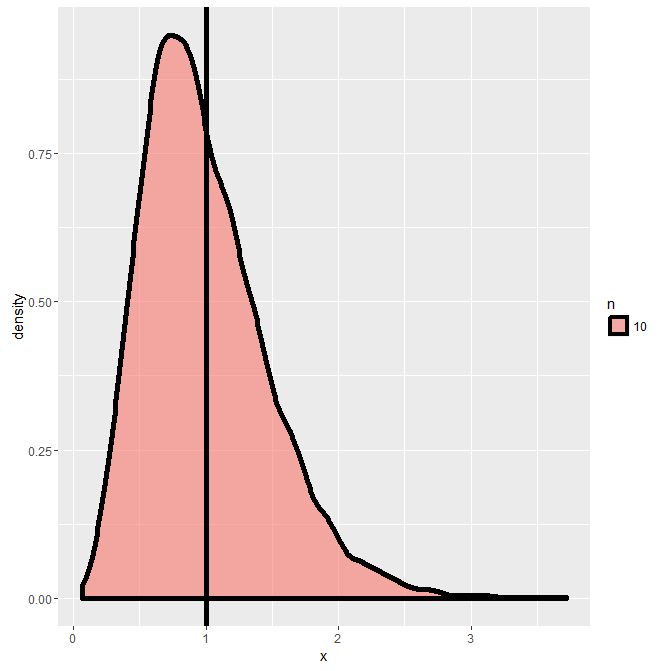
| As with the sample mean, the sample variance is also a random variable with an associated

| population distribution. Its expected value or mean is the population variance and its

| distribution gets more concentrated around the population variance with more data. The

| sample standard deviation is the square root of the sample variance.

...

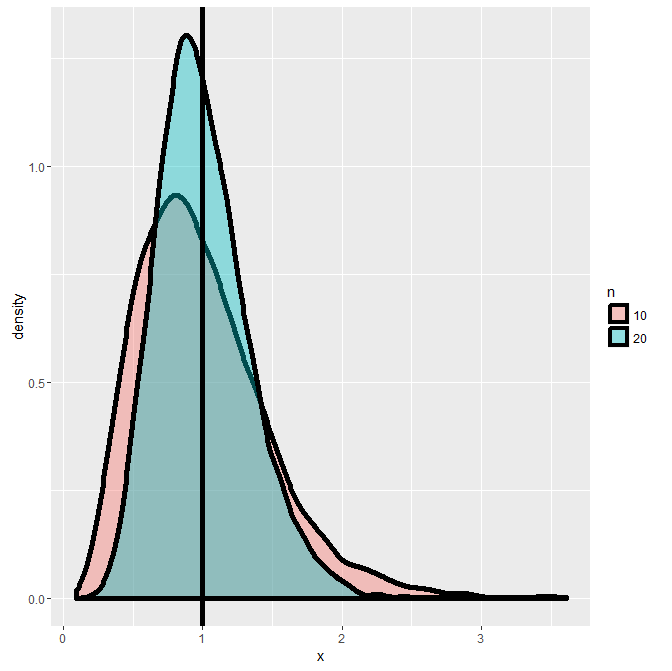


| |============================ | 35%

| To illustrate this point, consider this figure which plots the distribution of 10000

| variances, Each variance was computed on a sample of standard normals of size 10. The

| vertical line indicates the standard deviation 1.

... 

| |============================== | 37%

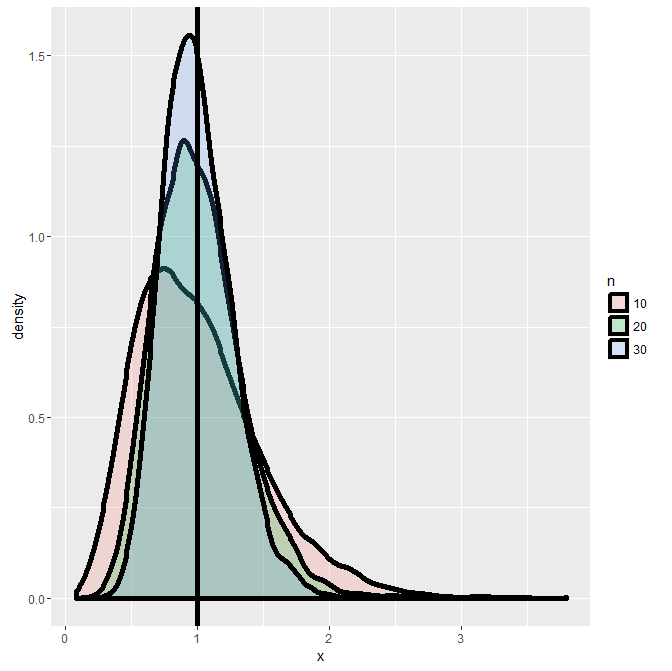
| Here we do the same experiment but this time (the taller lump) each of the 10000

| variances is over 20 standard normal samples. We've plotted over the first plot (the

| shorter lump) and you can see that the distribution of the variances is getting tighter

| and shifting closer to the vertical line.

...



| |================================ | 38%

| Finally, we repeat the experiment using 30 samples for each of the 10000 variances. You

| can see that with more data, the distribution gets more concentrated around the

| population variance it is trying to estimate.

...

| |================================= | 40%

| Now recall that the means of unbiased estimators equal the values they're trying to

| estimate. We can infer from the above that the sample variance is an unbiased estimator

| of population variance.

...

| |=================================== | 42%

| Recall that the average of random samples from a population is itself a random variable

| with a distribution centered around the population mean. Specifically, E(X') = mu, where

| X' represents a sample mean and mu is the population mean.

...

| |==================================== | 44%

| We can show that, if the population is infinite, the variance of the sample mean is the

| population variance divided by the sample size. Specifically, Var(X') = sigma^2 / n.

| Let's work through this in four short steps.

...

| |====================================== | 46%

| Which of the following does Var(X') equal? Here X' represents the sample mean and

| 'Sum(X\_i)' represents the sum of the n samples X\_1,...X\_n. Assume these samples are

| independent.

1: mu

2: sigma

3: E(1/n \* Sum(X\_i))

4: Var(1/n \* Sum(X\_i))

Selection: 2

| Not quite right, but keep trying.

| Which of the choices has both Var and the definition of mean in it?

1: mu

2: E(1/n \* Sum(X\_i))

3: sigma

4: Var(1/n \* Sum(X\_i))

Selection: 1

| Give it another try.

| Which of the choices has both Var and the definition of mean in it?

1: E(1/n \* Sum(X\_i))

2: mu

3: Var(1/n \* Sum(X\_i))

4: sigma

Selection: 1

| Almost! Try again.

| Which of the choices has both Var and the definition of mean in it?

1: mu

2: E(1/n \* Sum(X\_i))

3: Var(1/n \* Sum(X\_i))

4: sigma

Selection: 3

| Keep up the great work!

| |======================================= | 48%

| Which of the following does Var(1/n \* Sum(X\_i)) equal?

1: 1/n^2\*E(Sum(X\_i))

2: mu/n^2

3: sigma/n

4: 1/n^2\*Var(Sum(X\_i))

Selection: 3

| One more time. You can do it!

| Remember that fact about Var that we said would be useful before? Now is the time to use

| it.

1: mu/n^2

2: 1/n^2\*E(Sum(X\_i))

3: 1/n^2\*Var(Sum(X\_i))

4: sigma/n

Selection: 1

| That's not the answer I was looking for, but try again.

| Remember that fact about Var that we said would be useful before? Now is the time to use

| it.

1: 1/n^2\*Var(Sum(X\_i))

2: 1/n^2\*E(Sum(X\_i))

3: sigma/n

4: mu/n^2

Selection: 1

| That's a job well done!

| |========================================= | 50%

| Recall that Var is an expected value and expected values are linear. Also recall that our

| samples X\_1, X\_2,...,X\_n are independent. What does Var(Sum(X\_i)) equal?

1: E(mu)

2: Var(sigma)

3: E(Sum(X\_i))

4: Sum(Var(X\_i))

Selection: 4

| Great job!

| |=========================================== | 52%

| Finally, each X\_i comes from a population with variance sigma^2. What does Sum(Var(X\_i))

| equal? As before, Sum is taken over n values.

1: n\*(sigma)^2

2: n\*E(mu)

3: (n^2)\*Var(sigma)

4: n\*mu

Selection: 3

| Almost! Try again.

| Var(X\_i) is the constant value sigma^2 and we're summing over n of them.

1: (n^2)\*Var(sigma)

2: n\*mu

3: n\*(sigma)^2

4: n\*E(mu)

Selection: 3

| You got it right!

| |============================================ | 54%

| So we've shown that

| Var(X')=Var(1/n\*Sum(X\_i))=(1/n^2)\*Var(Sum(X\_i))=(1/n^2)\*Sum(sigma^2)=sigma^2/n for

| infinite populations when our samples are independent.

...

| |============================================== | 56%

| The standard deviation of a statistic is called its standard error, so the standard error

| of the sample mean is the square root of its variance.

...

| |=============================================== | 58%

| We just showed that the variance of a sample mean is sigma^2 / n and we estimate it with

| s^2 / n. It follows that its square root, s / sqrt(n), is the standard error of the

| sample mean.

...

| |================================================= | 60%

| The sample standard deviation, s, tells us how variable the population is, and s/sqrt(n),

| the standard error, tells us how much averages of random samples of size n from the

| population vary. Let's see this with some simulations.

...

| |================================================== | 62%

| The R function rnorm(n,mean,sd) generates n independent (hence uncorrelated) random

| normal samples with the specified mean and standard deviation. The defaults for the

| latter are mean 0 and standard deviation 1. Type the expression

| sd(apply(matrix(rnorm(10000),1000),1,mean)) at the prompt.

> sd(apply(matrix(rnorm(10000),1000),1,mean))

[1] 0.3133561

| Excellent job!

| |==================================================== | 63%

| This returns the standard deviation of 1000 averages, each of a sample of 10 random

| normal numbers with mean 0 and standard deviation 1. The theory tells us that the

| standard error, s/sqrt(n), of the sample means indicates how much averages of random

| samples of size n (in this case 10) vary. Now compute 1/sqrt(10) to see if it matches the

| standard deviation we just computed with our simulation.

> 1/sqrt(10)

[1] 0.3162278

| Keep working like that and you'll get there!

| |====================================================== | 65%

| Pretty close, right? Let's try a few more. Standard uniform distributions have variance

| 1/12. The theory tells us the standard error of means of independent samples of size n

| would have which standard error?

1: I haven't a clue

2: 1/sqrt(12\*n)

3: 1/(12\*sqrt(n))

4: 12/sqrt(n)

Selection: 2

| All that practice is paying off!

| |======================================================= | 67%

| Compute 1/sqrt(120). This would be the standard error of the means of uniform samples of

| size 10.

> 1/sqrt(120)

[1] 0.09128709

| Keep working like that and you'll get there!

| |========================================================= | 69%

| Now check it as we did before. Use the expression

| sd(apply(matrix(runif(10000),1000),1,mean)).

> sd(apply(matrix(runif(10000),1000),1,mean))

[1] 0.08776589

| Excellent job!

| |========================================================== | 71%

| Pretty close again, right? Poisson(4) are distributions with variance 4; what standard

| error would means of random samples of n Poisson(4) have?

1: I haven't a clue

2: 2\*sqrt(n)

3: 2/sqrt(n)

4: 1/sqrt(2\*n)

Selection: 4

| That's not exactly what I'm looking for. Try again.

| In this case s is 2. Divide this by sqrt(n).

1: 1/sqrt(2\*n)

2: 2/sqrt(n)

3: 2\*sqrt(n)

4: I haven't a clue

Selection: 2

| Excellent job!

| |============================================================ | 73%

| We'll do another simulation to test the theory. First, assume you're taking averages of

| 10 Poisson(4) samples and compute the standard error of these means. Use the formula you

| just chose.

> 2/sqrt(4)

[1] 1

| One more time. You can do it! Or, type info() for more options.

| Type '2/sqrt(10)' at the command prompt.

> 2/sqrt(10)

[1] 0.6324555

| Excellent job!

| |============================================================== | 75%

| Now check it as we did before. Use the expression

| sd(apply(matrix(rpois(10000,4),1000),1,mean)).

> sd(apply(matrix(rpois(10000,4),1000),1,mean))

[1] 0.6200033

| Nice work!

| |=============================================================== | 77%

| Like magic, right? One final test. Fair coin flips have variance 0.25; means of random

| samples of n coin flips have what standard error?

1: 2\*sqrt(n)

2: 2/sqrt(n)

3: 1/sqrt(2\*n)

4: I haven't a clue

5: 1/(2\*sqrt(n))

Selection: 2

| That's not exactly what I'm looking for. Try again.

| In this case s is 1/2 which is the sqrt of 1/4, the variance. Divide this by sqrt(n).

1: 1/sqrt(2\*n)

2: 2\*sqrt(n)

3: I haven't a clue

4: 1/(2\*sqrt(n))

5: 2/sqrt(n)

Selection: 1

| Nice try, but that's not exactly what I was hoping for. Try again.

| In this case s is 1/2 which is the sqrt of 1/4, the variance. Divide this by sqrt(n).

1: 2\*sqrt(n)

2: 1/sqrt(2\*n)

3: I haven't a clue

4: 2/sqrt(n)

5: 1/(2\*sqrt(n))

Selection: 5

| You are amazing!

| |================================================================= | 79%

| You know the drill. Assume you're taking averages of 10 coin flips and compute the

| standard error of these means with the theoretical formula you just picked.

> 1/(2\*sqrt(10))

[1] 0.1581139

| You are quite good my friend!

| |================================================================== | 81%

| Now check it as we did before. Use the expression

| sd(apply(matrix(sample(0:1,10000,TRUE),1000),1,mean)).

> sd(apply(matrix(sample(0:1,10000,TRUE),1000),1,mean))

[1] 0.1516859

| You got it right!

| |==================================================================== | 83%

| Finally, here's something interesting. Chebyshev's inequality helps interpret variances.

| It states that the probability that a random variable X is at least k standard deviations

| from its mean is less than 1/(k^2). In other words, the probability that X is at least 2

| standard deviations from the mean is less than 1/4, 3 standard deviations 1/9, 4 standard

| deviations 1/16, etc.

...

| |===================================================================== | 85%

| However this estimate is quite conservative for random variables that are normally

| distributed, that is, with bell-curve distributions. In these cases, the probability of

| being at least 2 standard deviations from the mean is about 5% (as compared to

| Chebyshev's upper bound of 25%) and the probability of being at least 3 standard

| deviations from the mean is roughly .2%.

...

| |======================================================================= | 87%

| Suppose you had a measurement that was 4 standard deviations from the distribution's

| mean. What would be the upper bound of the probability of this happening using

| Chebyshev's inequality?

1: 96%

2: 6%

3: 25%

4: 11%

5: 0%

Selection: 1

| Not quite, but you're learning! Try again.

| Chebyshev's inequality estimates that probability as 1/16. Convert this to a probability.

1: 11%

2: 96%

3: 6%

4: 0%

5: 25%

Selection: 3

| You are doing so well!

| |========================================================================= | 88%

| Now to review. The sample variance estimates what?

1: population variance

2: population

3: sample mean

4: sample standard deviation

Selection: 1

| All that hard work is paying off!

| |========================================================================== | 90%

| The distribution of the sample variance is centered at what?

1: population

2: sample mean

3: sample standard deviation

4: population variance

Selection: 2

| You're close...I can feel it! Try it again.

| What is the sample variance estimating?

1: population variance

2: population

3: sample standard deviation

4: sample mean

Selection: 1

| All that practice is paying off!

| |============================================================================ | 92%

| True or False - The sample variance gets more concentrated around the population variance

| with larger sample sizes

1: True

2: False

Selection: 1

| That's a job well done!

| |============================================================================= | 94%

| The variance of the sample mean is the population variance divided by ?

1: I haven't a clue

2: n

3: sqrt(n)

4: n^2

Selection: 3

| Not quite, but you're learning! Try again.

| Remember the 4 step proof starting with Var(X')=...? The last step had an n divided by an

| n^2.

1: n

2: n^2

3: I haven't a clue

4: sqrt(n)

Selection: 1

| Perseverance, that's the answer.

| |=============================================================================== | 96%

| The standard error of the sample mean is the sample standard deviation s divided by ?

1: n^2

2: n

3: I haven't a clue

4: sqrt(n)

Selection: 1

| Not exactly. Give it another go.

| Remember the many many examples we went through. The sqrt(n) figured prominently in them.

1: n

2: sqrt(n)

3: n^2

4: I haven't a clue

Selection: 2

| Your dedication is inspiring!

| |================================================================================ | 98%

| Congrats! You've concluded this vary long lesson on variance. We hope you liked it vary

| much.

...

| |==================================================================================| 100%

| Would you like to receive credit for completing this course on Coursera.org?

1: No

2: Yes

Selection: 2

What is your email address? sweeyean@gmail.com

What is your assignment token? I6Hx06c417dIgi50

Grade submission succeeded!

| You are amazing!

| You've reached the end of this lesson! Returning to the main menu...

| Please choose a course, or type 0 to exit swirl.

1: Statistical Inference

2: Take me to the swirl course repository!

Selection: