

Bootstrapping: how sure are your numbers?

You have five years of returns, a few dozen stores, a handful of deals. Every average, ratio or regression slope you compute from them would have come out differently with a different sample. The bootstrap measures how differently, using nothing but the data you have.

Functions `fx.RiskBootstrapIndex` · `fx.RiskBootstrap` · `fx.RiskBootstrapCI` · `fx.RiskBootstrapRegress`

In XL Edge XL Edge tab > Productivity Tools > Stat Tools > Insert Bootstrap

Example *XL Edge Example - Bootstrap.xlsx* (same folder). Every number in this guide is read from it.

Needs Excel for Microsoft 365. The functions also work without XL Edge once they are in a workbook.

- | | |
|---------------------------------------|--|
| 1 The idea in one paragraph | 6 Confidence intervals: BootstrapCI |
| 2 When to use it, and when not | 7 Regression standard errors: BootstrapRegress |
| 3 Using the ribbon | 8 Time series, shuffles and other options |
| 4 What a resample is: BootstrapIndex | 9 Pitfalls and good practice |
| 5 A bootstrap distribution: Bootstrap | 10 Argument reference |

1 The idea in one paragraph

Treat your sample as the best picture you have of the world it came from. Draw a new sample of the same size *from your sample*, picking rows at random **with replacement**, so some rows appear twice and others not at all, and compute your statistic on it. Do that a thousand times. The thousand results scatter around the original value, and **that scatter is the uncertainty of your statistic**: its standard error is their standard deviation, and a 95% confidence interval is (roughly) their middle 95%. No formula for the standard error is needed, which is the point. It works the same way for a median, a Sharpe ratio, a percentile or a regression slope, where textbook formulas are missing or rest on assumptions your data do not meet.

2 When to use it, and when not

Good uses	Use with care, or not at all
- Standard errors and intervals for statistics with no simple formula: median, percentiles, ratios, Sharpe ratio, correlation. - Checking textbook intervals when the data are skewed or fat-tailed, as returns usually are. - Regression standard errors when the scatter is uneven (one end noisier than the other). - Value at Risk from history, without assuming a normal distribution.	Very small samples (under about 20 rows): the sample is a poor picture of the world, so its resamples are too. Time series with momentum or seasonality: resampling single rows destroys the order. Use blocks (section 8). Extremes: the maximum, or a 99.9th percentile, can never exceed what is in the data. - It measures <i>sampling</i> uncertainty only. It cannot fix a biased sample or a wrong model.

3 Using the ribbon

XL Edge tab > Productivity Tools > Stat Tools (arrow) > Insert Bootstrap

The menu has three items: **Bootstrap Distribution**, **Bootstrap Confidence Interval** and **Bootstrap Regression**. Each walks you through a few short questions and writes one live formula:

- 1 Select the data:** a table with rows as observations (one or more columns), or any cell of a spilled result.
- 2 Choose the statistic:** type `mean`, `median`, `stdev`, a percentile such as `p5`, `return` (last ÷ first – 1 of the first column), or any formula of `s`, the resampled data, for example `AVERAGE(s)/STDEV.S(s)`.
- 3 Set the options:** number of resamples, and how rows are drawn (resample, blocks or shuffle) or which interval.
- 4 Pick where the result goes.** XL Edge tells you the block's size first and refuses to overwrite anything. Bootstrap Distribution also writes the histogram data and draws the chart beside it.

The result is an ordinary formula. Change the data and it recalculates. Edit the formula to change any choice. If the workbook does not yet have the functions, XL Edge installs the ones it needs first.

REPEATABLE BY DESIGN

The random draws come from a counter-based generator (HDR), so the same formula always gives the same numbers and editing an unrelated cell does not reshuffle them. For a different set of draws, change `Seed` (or `VarID`). Two bootstraps with the same Trials, VarID and Seed on data with the same number of rows use the same resamples.

4 What a resample is: `fx.RiskBootstrapIndexA`

This is the engine behind the other three, exposed so you can see it, or build your own bootstrap with ordinary formulas. It returns *row numbers*: one resample per row of output, each a draw of `N` numbers from 1...N.

```

5 resamples of the rows 1 ... 8 (in A6)
=fx.RiskBootstrapIndexλ(8, 5)
the deal sizes those rows pick (in A15)
=INDEX(Data!L5:L12, A6#)
each resample's average
=BYROW(A15#, LAMBDA(row, AVERAGE(row)))

```

On sheet *1 Resampling* the eight deals average **141.25** (\$ thousands). Each resample's average lands somewhere else; repeated a thousand times, the spread of those averages is the bootstrap standard error of the average. That is all the other functions do, at scale and with the statistics done for you.

Options. `BlockSize` > 1 draws runs of consecutive rows (for time series); `Replace = FALSE` gives permutations (every row exactly once, in a shuffled order) instead of resamples.

5 A bootstrap distribution: `fx.RiskBootstrapλ`

Applies your statistic to each resample and returns one value per resample, a column of `Trials` values, ready for a histogram, percentiles, or any summary. The example asks how reliable a fund's Sharpe ratio is, estimated from 60 months of returns:

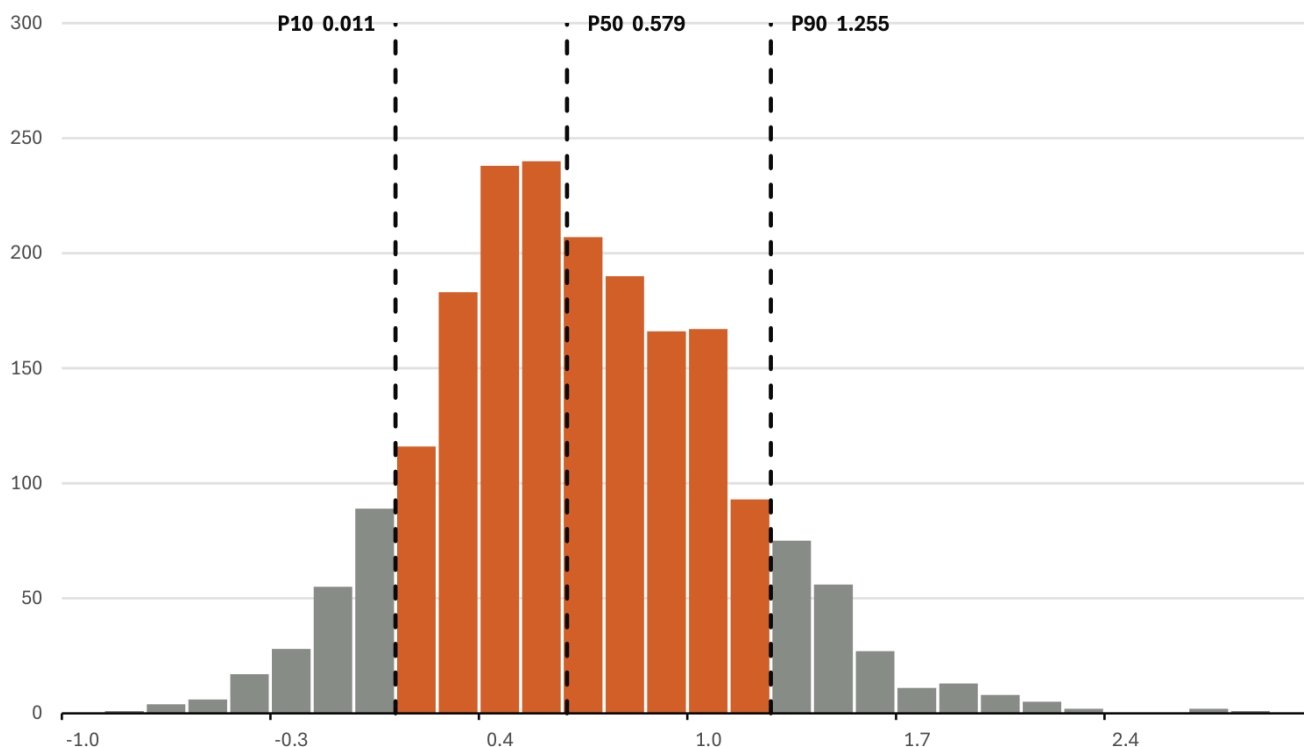
```

=fx.RiskBootstrapλ(Data!B5:B64, LAMBDA(s, AVERAGE(s) / STDEV.S(s) * SQRT(12)), 2000)

```

The `Statistic` argument is a `LAMBDA` of one argument (call it `s`) that takes a resample of `Data`, with the same shape, and returns one number. Anything Excel can compute from `s` works.

Bootstrap Sharpe ratio



The bootstrap distribution of the Sharpe ratio: 2,000 resamples, XL Edge's Bootstrap Distribution chart (orange = between the P10 and P90 lines). Sheet 2 *Distribution*.

Reading it. The data's own Sharpe ratio is **0.576**. The resamples average 0.617 with a standard deviation of **0.492**, which is the bootstrap standard error of the Sharpe ratio. Eight in ten land between 0.011 (P10) and 1.255 (P90), and **9.5%** are below zero. Five years of data cannot pin this fund's risk-adjusted return down much better than "somewhere between poor and very good", a common and humbling result for Sharpe ratios.

SUMMARIES THAT WORK ON THE OUTPUT

`=STDEV.S(A6#)` standard error · `=PERCENTILE.INC(A6#, {0.05, 0.95})` a quick percentile interval ·
`=AVERAGE(--(A6# < 0))` share below zero · `=fx.RiskStatsλ(A6#)` the full summary block ·
`=fx.RiskVaλ(A6#, 0.99)` when the statistic is a return or P&L.

6 Confidence intervals: `fx.RiskBootstrapCIA`

Turns a bootstrap into an interval in one step. It returns a 12×2 block: the estimate, the bootstrap mean, the **bias** (bootstrap mean – estimate), the standard error, the interval's lower and upper ends, the percentiles those ends sit at (handy for marking them on the histogram), the method and the number of resamples.

```
BCa (the default), 95% (the default), 2,000 resamples
=fx.RiskBootstrapCIA(Data!B5:B64, LAMBDA(s, AVERAGE(s)), , , 2000, , , , "Mean return")
the other two methods, same draws
=fx.RiskBootstrapCIA(Data!B5:B64, LAMBDA(s, AVERAGE(s)), , , "Percentile", 2000)
=fx.RiskBootstrapCIA(Data!B5:B64, LAMBDA(s, AVERAGE(s)), , , "Basic", 2000)
```

95% intervals for the fund's average monthly return (estimate 0.619%), sheet 3 *Interval*

Method	Lower	Upper	How it is built
BCa (default)	-0.347%	1.520%	Percentiles shifted for the bootstrap's median bias and skewness (acceleration from a jackknife). The most accurate of the three.
Percentile	-0.335%	1.538%	The 2.5th and 97.5th percentiles of the bootstrap values. Simple; fine when the distribution is near symmetric.
Basic	-0.300%	1.573%	2 × estimate minus those percentiles, swapped. Corrects for a shifted center, but not for skewness.
t-interval (textbook)	-0.343%	1.581%	Mean ± t × SD / √n, for comparison.

For a mean of 60 months the four agree closely, as they should; the bootstrap standard error (0.475%) is the textbook one in disguise. The bootstrap earns its keep where no textbook formula exists. The fund's **median** monthly return is 1.355%, and its BCa interval runs from **-0.095%** to **1.760%**. Notice that it is lopsided around the estimate, which a "± something" formula could never show.

HOW MANY RESAMPLES?

1,000 is plenty for a standard error. For interval ends, especially BCa's, use 2,000 or more; for 99% intervals, 5,000 or more. If two Seeds give visibly different ends, add resamples.

7 Regression standard errors: `fx.RiskBootstrapRegressA`

Bootstraps a linear regression's coefficients. It returns, for each coefficient, the OLS estimate, a **bootstrap standard error** and a percentile interval: an (m + 2) × 5 block, where m is the number of coefficients. There are two ways to resample:

- **Pairs** (default) resamples whole rows (y with its x's). It makes no assumption about how the scatter around the line behaves, so it stays right when the scatter is uneven or the straight line is only approximate.
- **Residuals** keeps the x's fixed and resamples the residuals onto the fitted line (rescaled for leverage). It is more precise *if* the scatter really is the same everywhere, which is exactly what OLS assumes.

```
ordinary least squares, for comparison
=fx.RiskRegressA(Data!H5:H52, Data!F5:G52, , , {"Units","Price","Promotion"})
bootstrap standard errors: pairs (the default), then residuals
=fx.RiskBootstrapRegressA(Data!H5:H52, Data!F5:G52, , , 2000)
=fx.RiskBootstrapRegressA(Data!H5:H52, Data!F5:G52, "Residuals", , 2000)
```

Units sold on price and promotion spend, 48 stores, sheet 4 *Regression*

Coefficient	Estimate	OLS SE	Residuals SE	Pairs SE
Price	-24.01	3.34	3.59	3.08
Promotion	17.79	2.67	2.81	6.36

In these stores, most run small promotions and a few run big ones, and sales scatter far more where the promotion is big. OLS, and the residual bootstrap, which shares its assumption, put the promotion coefficient's standard error near 2.67. The pairs bootstrap, which sees the uneven scatter, puts it at **6.36**, about 2.4 times larger, and its 95% interval runs from 7.27 to 30.35 extra units per \$1,000 of promotion. The effect is real but far less precisely known than OLS claims. For price, whose effect on the scatter is small, the three are similar (3.08 to 3.59): no sign of the problem promotion has.

A QUICK DIAGNOSIS

If the pairs standard errors are much larger than the OLS ones, the scatter is uneven (heteroskedastic) and the OLS p-values are too optimistic. XL Edge's Breusch-Pagan / White test (Insert Statistical Diagnostics) gives the formal answer.

8 Time series, shuffles and other options

Blocks, for data with memory

Resampling single rows assumes the rows are independent. Monthly returns are nearly so; sales, prices and most operating series are not. They trend, cycle and carry momentum. `BlockSize` draws runs of consecutive rows (a moving-block bootstrap), which keeps short-range dependence inside each block. Pick a block length of a few times the dependence you expect; for monthly data, 6 to 12 is a common start. On the example fund, 6-month blocks give a Sharpe standard error of 0.471 against 0.492 for single months. They are close, confirming that its returns carry little memory.

Shuffles (permutations)

`Replace = FALSE` uses every row exactly once, in a random order. That is useless for a mean (every shuffle has the same mean) but right for *order-dependent* statistics such as a path's return, last ÷ first – 1. With several columns, `Columns = "Independent"` shuffles each column on its own, breaking the links between them. That is how Benninga (*Financial Modeling*, ch. 28.5) builds a historical-simulation Value at Risk: shuffle each risk factor's history independently, revalue the position along each shuffled path, take the return, and read the VaR from those returns with `fx.RiskVaR`. The ribbon offers this as the `shuffle` choice with the `return` statistic.

Several columns at once

`Data` may have several columns. By default (`Columns = "Together"`) whole rows are drawn, so the columns keep their relationship, which is what a correlation or a portfolio statistic needs. Inside the LAMBDA, take columns with `CHOOSECOLS(s, 1)`, for example `LAMBDA(s, CORREL(CHOOSECOLS(s,1), CHOOSECOLS(s,2)))`.

9 Pitfalls and good practice

- **Garbage in, garbage out.** The bootstrap knows only your sample. A period without a crisis cannot tell you about crises.
- **Small samples understate uncertainty.** With 15 rows, intervals come out too narrow. Report them with a caveat, or not at all.
- **Watch for dependence.** If the series has memory, single-row resamples give standard errors that are too small. Use blocks.
- **Don't bootstrap extremes.** Minimums, maximums and far tails are bounded by the data. Use a fitted distribution instead.
- **Bias is information.** A large bias (bootstrap mean far from the estimate) warns that the statistic is skewed or unstable. BCa corrects for it; the percentile method does not.
- **Keep the seed visible.** Write Trials, VarID and Seed into the formula so anyone can reproduce your numbers.
- **Speed.** Cost grows with $\text{Trials} \times \text{rows} \times \text{the statistic's own cost}$. A few thousand resamples of a few hundred rows calculate in seconds. BCa adds one evaluation per row (the jackknife).

10 Argument reference

```
fx.RiskBootstrapIndexλ(N, [Trials], [BlockSize], [Replace], [VarID], [Seed])
```

Argument	Default	Meaning
N	required	Number of rows in the data.
Trials	1000	Resamples (rows of output). Output: Trials × N row numbers.
BlockSize	1	Length of the runs of consecutive rows drawn (moving-block bootstrap).
Replace	TRUE	FALSE for permutations: each row exactly once, shuffled.
VarID, Seed	1, 0	Select the random stream. Change either for new draws.

```
fx.RiskBootstrapλ(Data, Statistic, [Trials], [BlockSize], [Replace], [Columns], [VarID], [Seed])
```

Argument	Default	Meaning
Data	required	n × k, rows are observations.
Statistic	required	LAMBDA(s, ...) returning one number from a resample s of Data.
Trials	1000	Resamples. Output: Trials × 1.
BlockSize, Replace	1, TRUE	As for BootstrapIndex.
Columns	"Together"	"Independent" draws each column on its own (Benninga's shuffle).
VarID, Seed	1, 0	The random stream.

```
fx.RiskBootstrapCIAλ(Data, Statistic, [Level], [Method], [Trials], [BlockSize], [VarID], [Seed], [Name])
```

Argument	Default	Meaning
Level	0.95	Confidence level of the interval.
Method	"BCa"	"BCa", "Percentile" or "Basic" (section 6).
Trials	1000	Use 2,000+ for interval ends.
Name	"Statistic"	Shown in the block's title.

Output, 12 × 2: title; Item / Value; Estimate; Bootstrap mean; Bias; Std error; Lower; Upper; Lower and Upper percentile used; Method; Trials.

```
fx.RiskBootstrapRegressλ(Y, X, [Method], [Level], [Trials], [VarID], [Seed], [Intercept])
```

Argument	Default	Meaning
Y, X	required	$n \times 1$ dependent variable; $n \times k$ explanatory variables.
Method	"Pairs"	"Pairs" (whole rows) or "Residuals" (fixed X, resampled residuals).
Level, Trials	0.95, 1000	Interval level and resamples.
Intercept	TRUE	FALSE to fit through the origin.
Output, $(m + 2) \times 5$: title; Term / Estimate / Bootstrap SE / Lower / Upper; one row per coefficient.		

References: Efron (1979), *Annals of Statistics* 7(1); Efron and Tibshirani (1993), *An Introduction to the Bootstrap*; Davison and Hinkley (1997), *Bootstrap Methods and their Application*; Künsch (1989), *Annals of Statistics* 17(3); Benninga, *Financial Modeling*, 4th ed., ch. 28. The example data are simulated for illustration. XL Edge and the fx.Risk... functions: © 2026 W Phillips, edgewisedata.com, MIT License.