Enveritas

How many coffee farms are there in the world?

David Browning

ICO Kenya

March 27 2019

25 million coffee farmers is a common global estimate

Google

how many coffee farmers are there in the world

All News Images Shopping Videos More Settings Tools

About 174,000,000 results (0.56 seconds)

Coffee farmers. Coffee is one of the world's most popular beverages and **80%** of it is produced by **25 million** smallholders. Around **125 million** people worldwide depend on coffee for their livelihoods. It is the most valuable and widely traded tropical agricultural product and is mainly produced by smallholder farmers.

The New York Times

"An estimated **25 million** of these farmers have suffered..." [1]

INTERNATIONAL COFFEE ORGANIZATION

"**25 million** coffee farmers are dependent on governments, companies, coffee cooperatives, trades unions and NGOs coming together to solve the problem." [4]

PBS

"Nearly **25 million** farmers worldwide depend on growing coffee for their economic livelihood." [3]

IISD
International Institute for Sustainable Development

"Estimates of total coffee farmers worldwide have long hovered at about **20 million to 25 million**." [2]

[1] https://www.nytimes.com/2002/11/03/business/business-global-issues-flow-into-america-s-coffee.html [2] https://www.iisd.org/pdf/2014/ssi_2014_chapter_8.pdf [3] http://www.pbs.org/frontlineworld/stories/guatemala/mexicofacts.html [4] https://www.oxfamamerica.org/publications/mugged-poverty-in-your-coffee-cup/

To the best of our knowledge, there is no published source of the data or methodology for this frequently cited number

We interviewed more than **20,000 farm households**

and consulted the expertise of over **80 institutions**

then applied **statistical models**

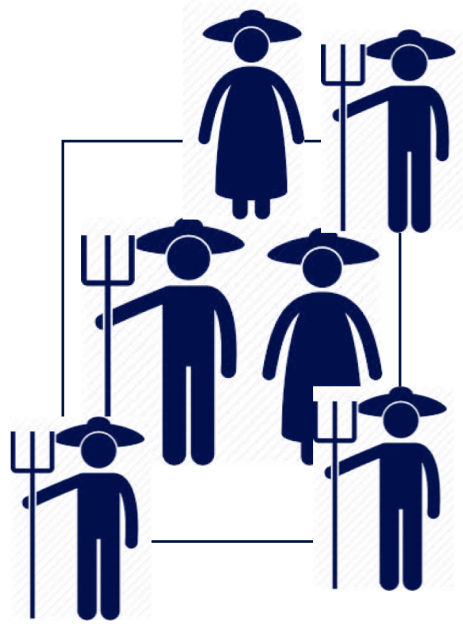
to understand **productivity & farm size** characteristics

To develop an estimate of global coffee farms


Kahawatu
INTERNATIONAL COFFEE ORGANIZATION
NCE
NAIROBI COFFEE EXCHANGE
SUCAFINA
Rainforest Alliance
IITA
Transforming African Agriculture
Hivos
people unlimited
COSA
Committee On Sustainability Assessment
Central Statistical Agency
Addis Ababa, Ethiopia
Oromia Coffee Farmers Cooperative Societies Union (OCFCSU) LTD.
Coffee Management Services
THE COFFEE MANAGEMENT EXPERTS
TANZANIA COFFEE BOARD
Uganda Coffee Development Authority
UCFA
ROYAL PACIFIC INDUSTRY
BEYOND RELIABLE
P.T. TOARCO JAYA
Icafe
Instituto del Café de Costa Rica
ECX
ETHIOPIA COMMODITY EXCHANGE
Kenya Coffee Traders Association
K.C.T.A
TaCRI
TANZANIA COFFEE RESEARCH INSTITUTE
Uganda Bureau of Statistics
UGACOF
Coricafé S.A.
SIDAMA COFFEE
FARMERS COOPERATIVE UNION
CPT
Coffee Partnership for Tanzania
Café Africa
HANI COFFEE Co
哈美尼咖啡
AYO INDONESIA
Escuela de Café de El Salvador
TECHNOSERVE
BUSINESS SOLUTIONS TO POVERTY
NAEB
Hanns R. Neumann Stiftung
KYAGALANYI COFFEE LTD
Member of VCLCAFE Group
YUNNAN COFFEE TRADERS
rootcapital
WORLD COFFEE RESEARCH
COFFEE GORILLA SHOP
VOLCAFE
GLOBAL COFFEE PLATFORM
for a sustainable coffee world
Lutheran World Relief
Anacafé
ASOCIACION NACIONAL DEL CAFE
INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI
Yirgacheffe Coffee Farmers Cooperatives Union
MSU RESEARCH
KCU
QUALITY COFFEE
Kagera Cooperative Union
Metajua
Collect - Analyze - Display
COFFEE BOARD
1958-2008
Agriculture and Food Authority (AFA)
Coffee Directorate
Our Crops, Our Wealth
Rogers Family Company
Olam
COFFEE LAB INTERNATIONAL
IDH
IFC
International Finance Corporation
World Bank Group
cooxupé
LDC.
Louis Dreyfus Company

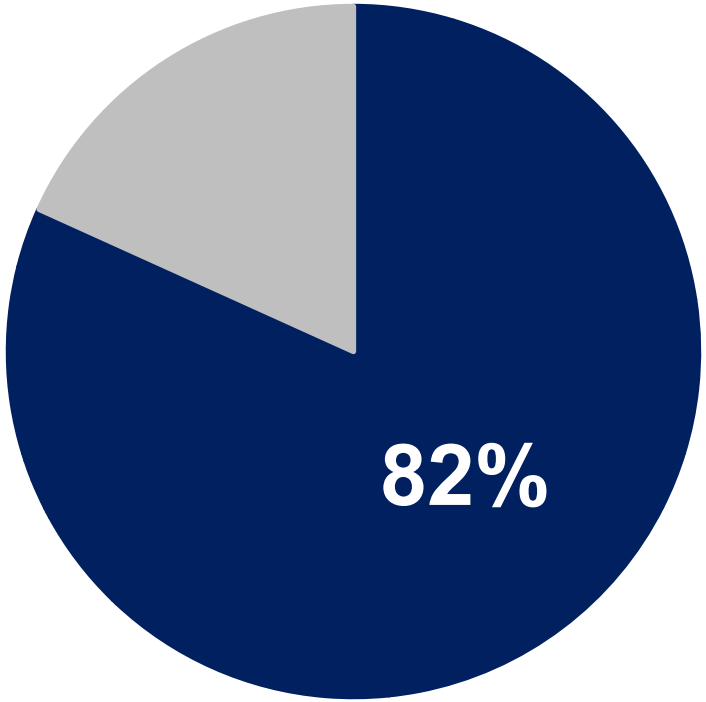
First a caveat, this is not a perfect number

Second, a few definitions to avoid ambiguity

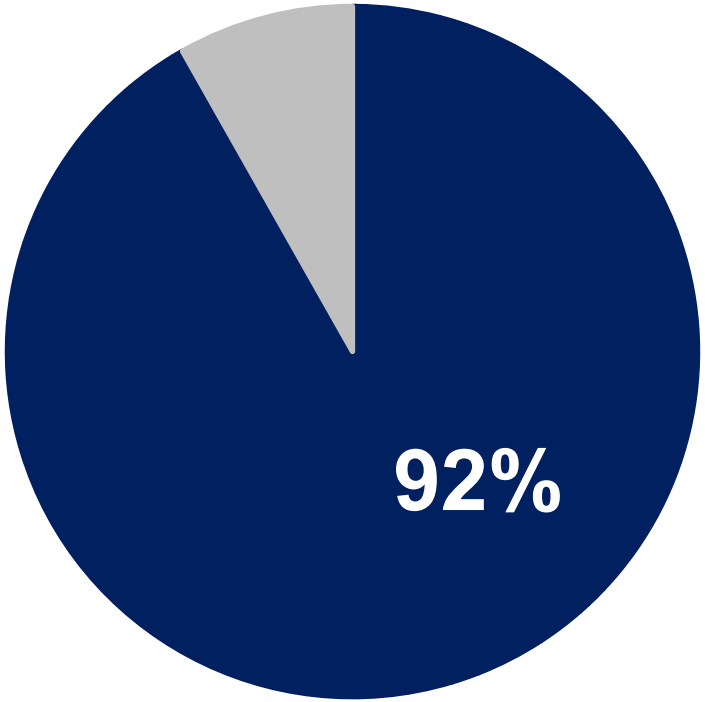


We excluded coffee farms that produce only for own consumption where it would distort the data

Our analysis covered origins with most of the worlds coffee farmers and coffee volume

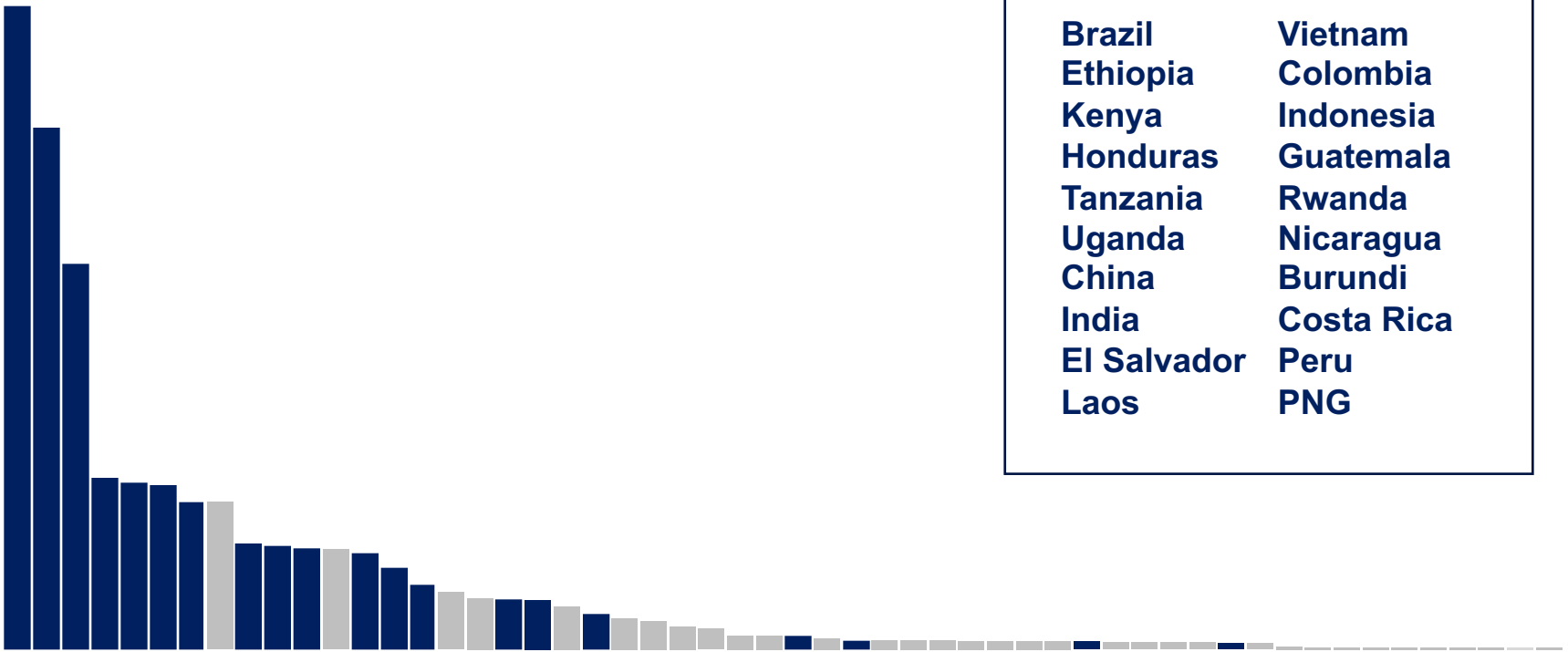


% of farm population



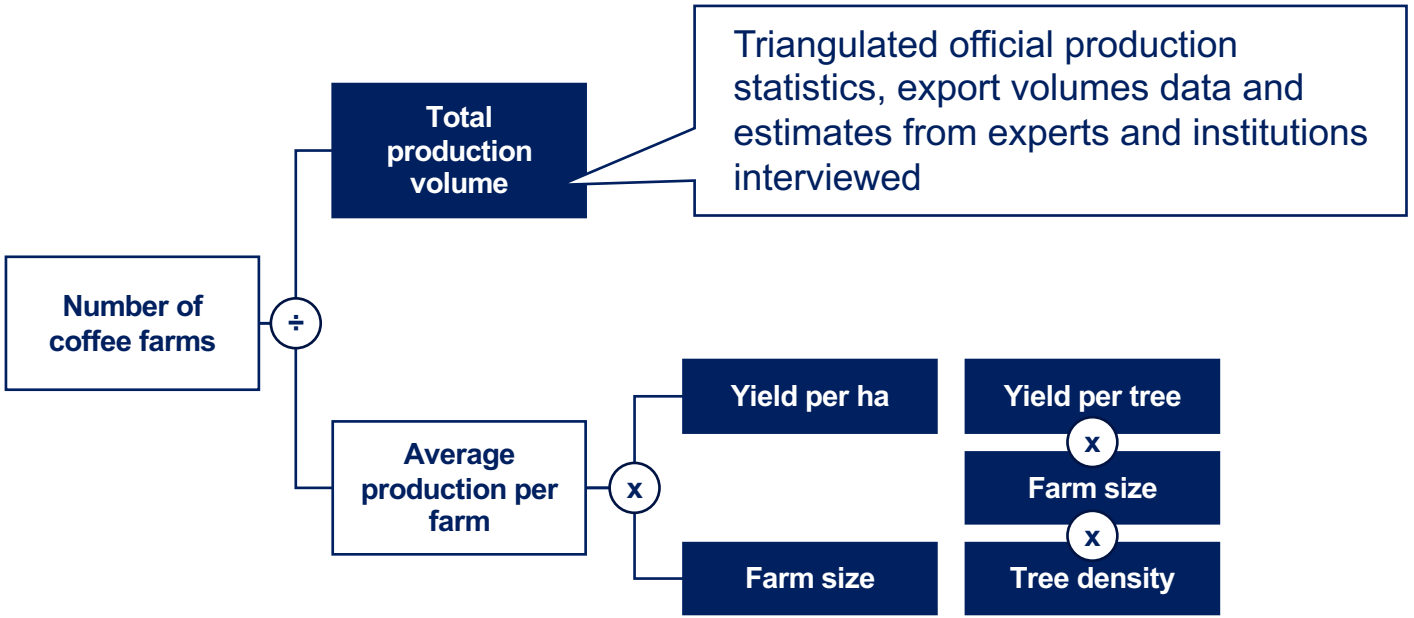
% of global production

Analyzed 20 countries of the main 58 origins



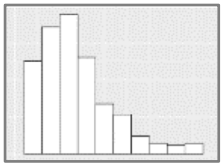
Brazil	Vietnam
Ethiopia	Colombia
Kenya	Indonesia
Honduras	Guatemala
Tanzania	Rwanda
Uganda	Nicaragua
China	Burundi
India	Costa Rica
El Salvador	Peru
Laos	PNG

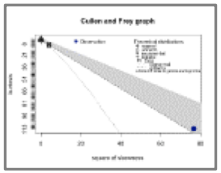
How did we estimate the number of coffee farms per country?

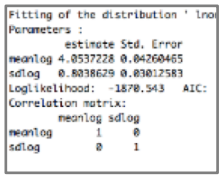


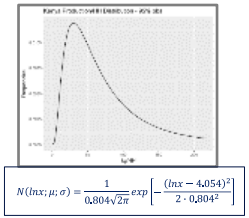
We take sample data through 4 steps to produce a population average

	1 Plot histogram of sample data	2 Decide on population distribution	3 Estimate distribution parameters	4 Plot distribution and estimate population average
Description:	• Plot the histogram of the sample data • Have a first look at the distribution of data	• Decide on which distribution type 'mimics' best the sample data by using the Cullen and Frey graph • Plot the theoretical and empirical densities for selected distributions to confirm the distribution choice • Test if the data follows chosen distribution	• After having selected the distribution curve, estimate the parameters of the distribution (shape and scale)	• Plot the distribution curve • Calculate the average(median) of the population using the equation of the distribution curve • Plot the Cumulative Distribution Function to determine the quartile distribution
Output:	• Histogram of sample data	• Chosen distribution curve	• Equation of distribution curve	• Population median

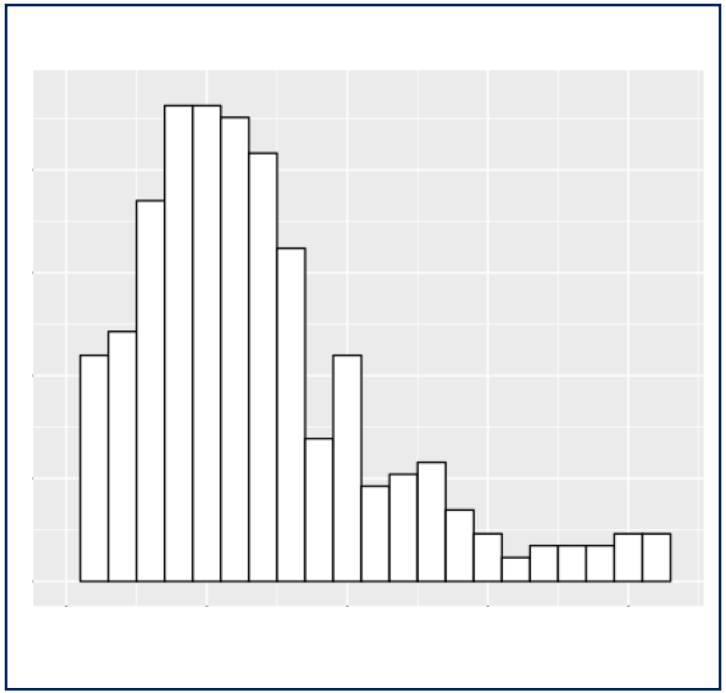








1 Plot the histogram of the sample data

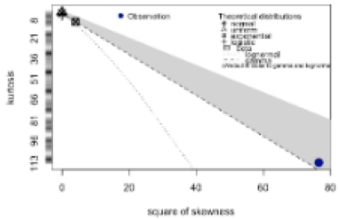


- Histogram gives a rough sense of the density of the underlying distribution of the data, and often for density estimation: it helps to get the sense of what the distribution of the population might be.
- The total area of a histogram is always equal to 1.

2 Decide on the population distribution

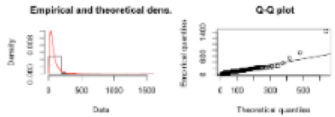
Cullen & Frey graph

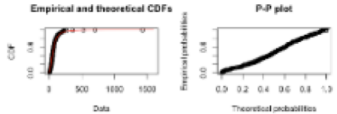
- Use the Cullen and Frey graph to decide on the distribution



Empirical and theoretical densities

- Plot the theoretical and empirical densities for chosen distributions
- Confirm the distribution choice





Hypothesis testing

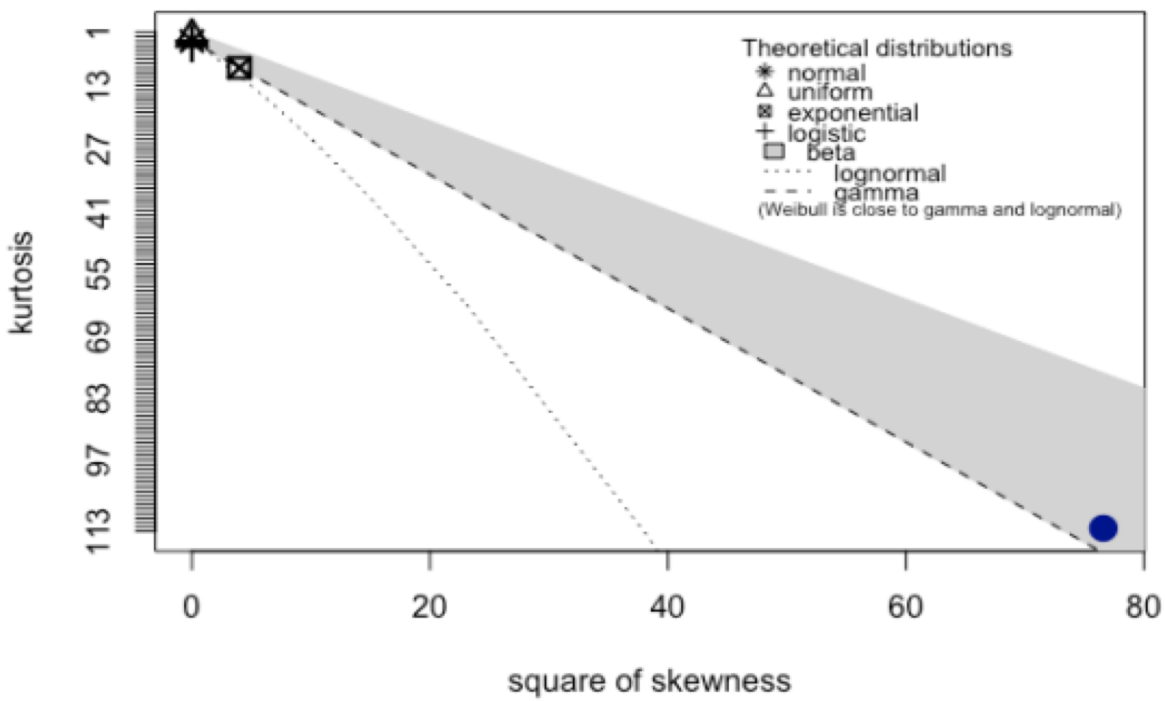
- Hypothesis testing to see if the data follows chosen distribution

```
Goodness-of-fit statistics
                                 lnorm     gamma
Kolmogorov-Smirnov statistic 0.07702247 0.0935085
Cramer-von Mises statistic   0.49492226 0.7335974
Anderson-Darling statistic   2.91063191 3.9774881

Goodness-of-fit criteria
                                 lnorm    gamma
Akaike's Information Criterion 3745.086 3777.756
Bayesian Information Criterion 3752.836 3785.506
```

2 Use the Cullen and Frey graph to recognize the possible distribution of population from which the sample is drawn

- **Cullen & Frey graph**
- Empirical and theoretical densities
- Hypothesis testing



Cullen and Frey graph plots the observations from data set (blue dot) against various distributions.

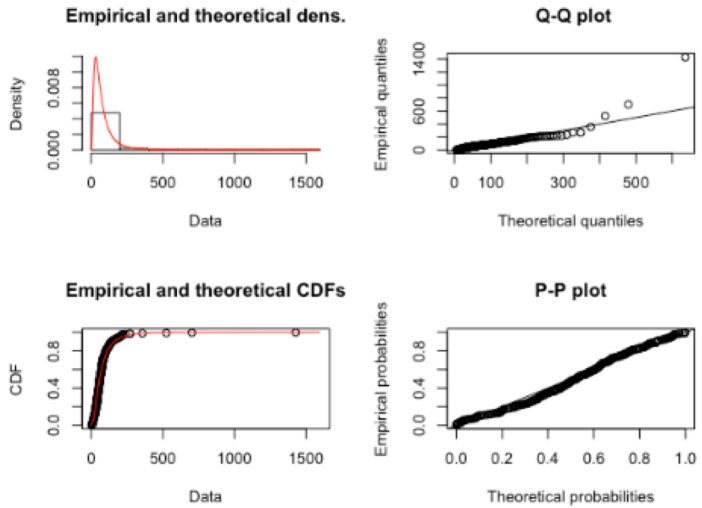
2 Compare our data with log normal and gamma theoretical distributions

Cullen & Frey graph

Empirical and theoretical densities

Hypothesis testing

- Plot the theoretical and empirical densities for chosen distributions
- Confirm the distribution choice



These plots help to determine if the data set come from population with given distribution (lognormal in this example). The theoretical values (represented by the solid line) against the empirical values (dots) are plotted.

2 Then we conduct two additional tests to check our distribution choice of log normal

Cullen & Frey graph

Empirical and theoretical densities

Hypothesis testing

- Hypothesis testing to see if the data follows chosen distribution

```
Goodness-of-fit statistics
                                 lnorm     gamma
Kolmogorov-Smirnov statistic 0.07702247 0.0935085
Cramer-von Mises statistic   0.49492226 0.7335974
Anderson-Darling statistic   2.91063191 3.9774881

Goodness-of-fit criteria
                                 lnorm    gamma
Akaike's Information Criterion 3745.086 3777.756
Bayesian Information Criterion 3752.836 3785.506
```

The Kolmogorov-Smirnov test is a nonparametric test of the equality of continuous, one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution. The null hypothesis is that the sample is drawn from the reference distribution.

Akaike's and Bayesian Information Criterion (AIC, BIC) are criterions for model selection among a finite set of models; the model with the lowest BIC/AIC is preferred.
AIC and BIC are strongly related to each other.

3 Once the distribution is chosen we need to estimate the parameters of the distribution

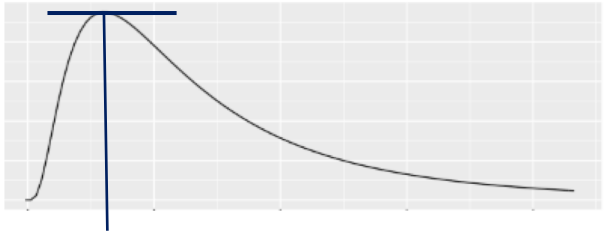
```
Fitting of the distribution ' lnorm ' by maximum likelihood
Parameters :
         estimate Std. Error
meanlog 4.0537228 0.04260465
sdlog   0.8038629 0.03012583
Loglikelihood:  -1870.543   AIC:  3745.086   BIC:  3752.836
Correlation matrix:
        meanlog sdlog
meanlog       1     0
sdlog         0     1
```

$$N(lnx;\mu;\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} exp\left[-\frac{(lnx-\mu)^2}{2\sigma^2}\right]$$

:Now need to estimate two parameters:
- meanlog (scale – average of the distribution)
- sdlog (shape – standard deviation of the log of the distribution)

4 Three choices (measures of central tendency) to provide an average for the meanlog

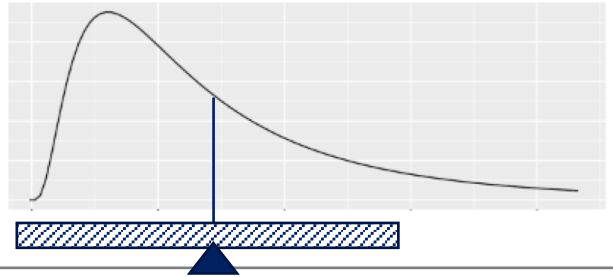
Mode



The **MODE** refers to the value that appears most often.

Often used for non-numerical variables as for numerical variables usually mean and median are of better choice.

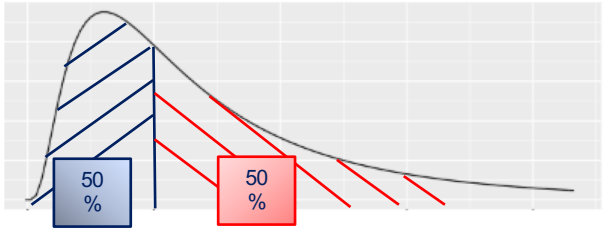
Mean



The **MEAN** refers to the central value of the dataset. It is determined by adding all the data points in a population and then dividing the total by the number of points.

Most popular measure of central tendency but not the best choice when the distribution is skewed as it is highly influenced by outliers.

Median

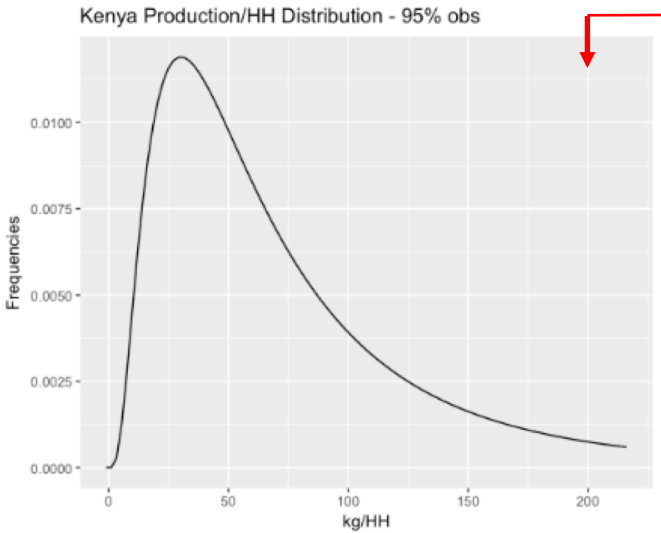


The **MEDIAN** refers to **the middle value of the dataset**. It divides the dataset into two: 50% of values of the dataset are lower to the median and 50% of values are higher than the median.

We have used the median as it is more robust for skewed distributions with long tail ends

4 Once we have the parameters we can plot the distribution of the population and create a function

Plotting distribution based on the estimates




```
Fitting of the distribution ' lnorm ' by maximum likelihood
Parameters :
         estimate Std. Error
meanlog 4.0537228 0.04260465
sdlog   0.8038629 0.03012583
Loglikelihood:  -1870.543   AIC:  3745.086   BIC:  3752.836
Correlation matrix:
        meanlog sdlog
meanlog       1     0
sdlog         0     1
```

+

$$N(lnx;\mu;\sigma) = \frac{1}{0.804\sqrt{2\pi}} exp\left[-\frac{(lnx-4.054)^2}{2\cdot 0.804^2}\right]$$

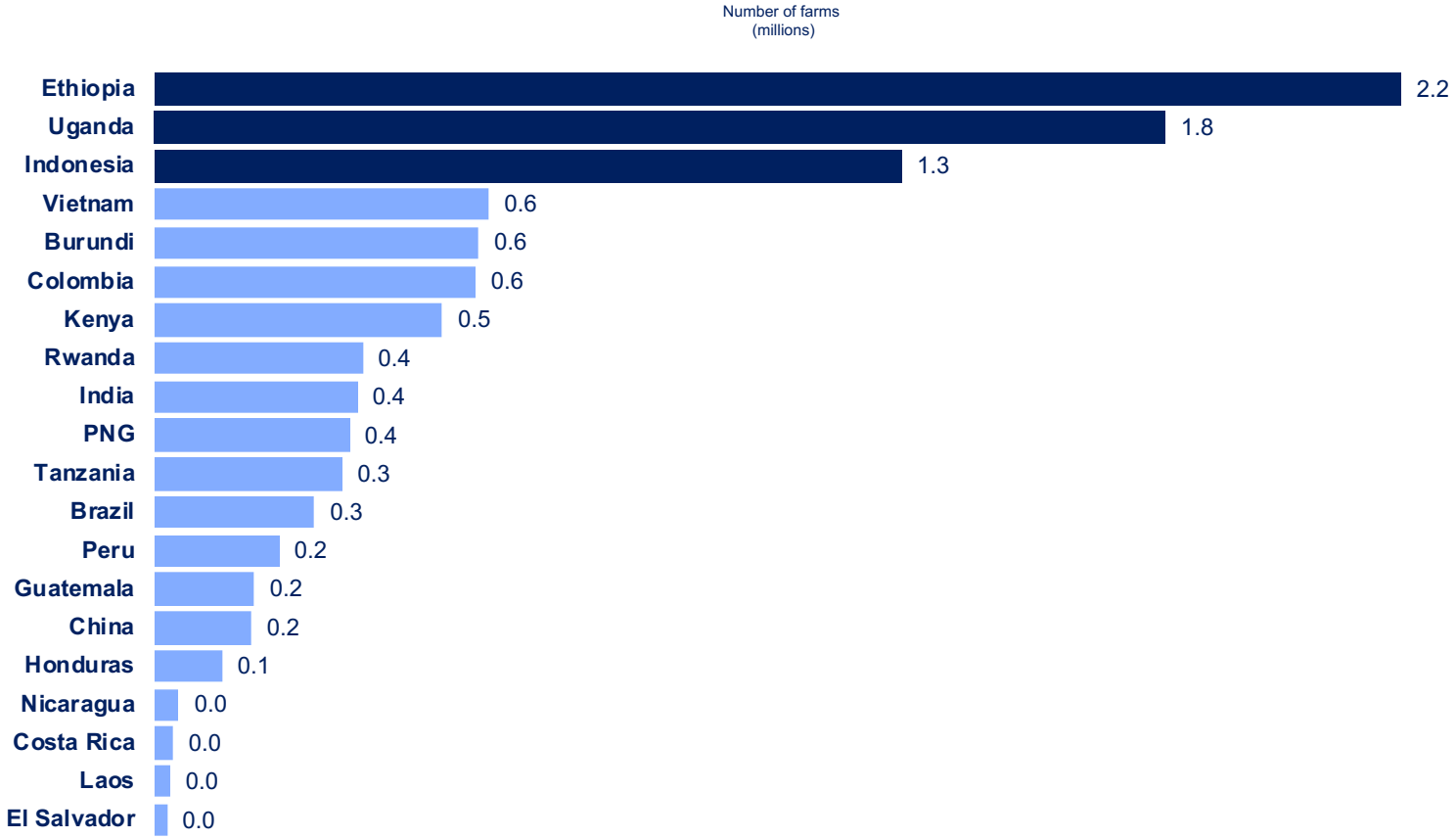
Estimating population median

Once distribution is chosen the parameters of the distribution are estimated and thus the population parameters can be calculated.

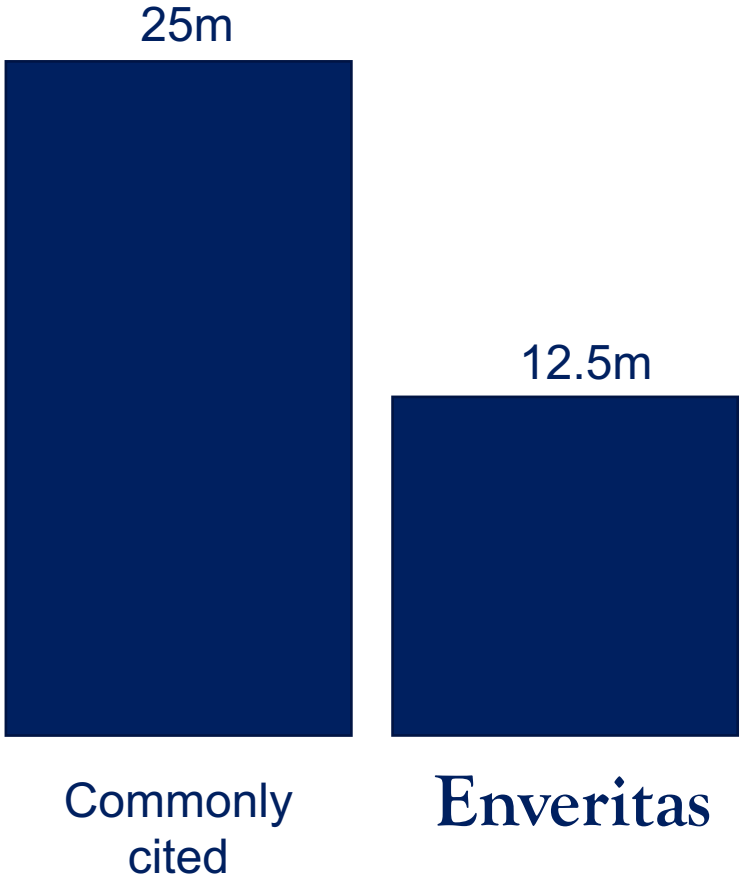
Median$= \exp(\mu)$

Median = 57.6 kg per farm

3 countries represent nearly half of the worlds coffee farms



We estimate 12.5 million coffee farms globally, half the more commonly cited number of 25 million



Thank you

- The many institutions that gave generously of their time
- 20,000 coffee farmers that patiently worked with us
- The Enveritas team