1C 2014 NNN2 1

# Exploring Lognormal Incomes

**Milo Schield**
**Augsburg College**
**Editor: www.StatLit.org**
**US Rep: International Statistical Literacy Project**

**11 October 2014**
**National Numeracy Network**

**www.StatLit.org/pdf/2014-Schield-NNN2-Slides.pdf**
**www.StatLit.org/Excel/Create-LogNormal-Incomes-Excel2013.xlsx**

1C 2014 NNN2 2

## Log-Normal Distributions

A Log-Normal distribution is generated from a normal with mu = Ln(Median) and sigma = Sqrt[2*Ln(Mean/Median)].

The lognormal is always positive and right-skewed.

Examples:
- Incomes (bottom 97%), assets, size of cities
- Weight and blood pressure of humans (by gender)

Benefit:
- calculate the share of total income held by the top X%
- calculate Gini Coefficient,
- explore effects of change in mean-median ratio.

1C 2014 NNN2 3

## Log-Normal Distributions

"In many ways, it [the Log-Normal] has remained the Cinderella of distributions, the interest of writers in the learned journals being curiously sporadic and that of the authors of statistical test-books but faintly aroused."

"We … state our belief that the lognormal is as fundamental a distribution in statistics as is the normal, despite the stigma of the derivative nature of its name."

Aitchison and Brown (1957). P 1.

1C 2014 NNN2 4

## Lognormal and Excel

Use Excel to focus on the model and the results.

Excel has two Log-Normal functions:

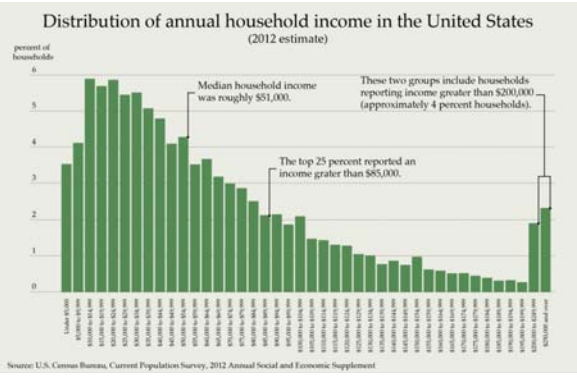
- Standard: =LOGNORM.DIST(X, mu, sigma, k)
  k=0 for PDF; k=1 for CDF.
- Inverse: =LOGNORM.INV(X, mu, sigma)

Use Standard to calculate/graph the PDF and CDF.

Use Inverse to find cutoffs: quartiles, to 1%, etc.

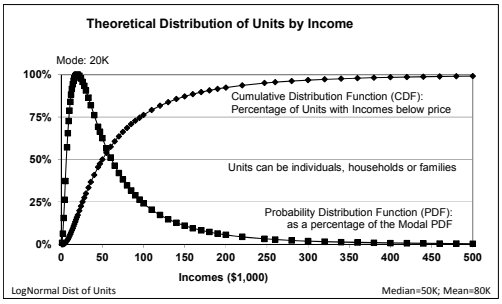
Use Excel to create graphs that show comparisons.

1C 2014 NNN2 5




1C 2014 NNN2 6

## Log-Normal Distribution of Units



1C 2014 NNN2 7

## Paired Distributions

For anything that is distributed by X, there are always two distributions:

1. Distribution of subjects by X
2. Distribution of total X by X.

Sometime we ignore the 2nd: height or weight.

Sometimes we care about the 2nd: income or assets.

Surprise: If the 1st is lognormal, so is the 2nd.

1C 2014 NNN2 8

## Distribution of Households and Total Income by Income

Suppose the distribution of households by income is log-normal with normal parameters mu# and sigma#.

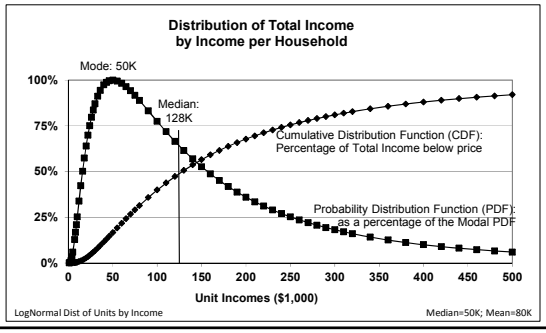
Then the distribution of total income by amount has a log-normal distribution with these parameters: mu$ = mu# + sigma#^2; sigma$ = sigma#.

See Aitchison and Brown (1963) p. 158.

Special thanks to Mohammod Irfan (Denver University) for his help on this topic.

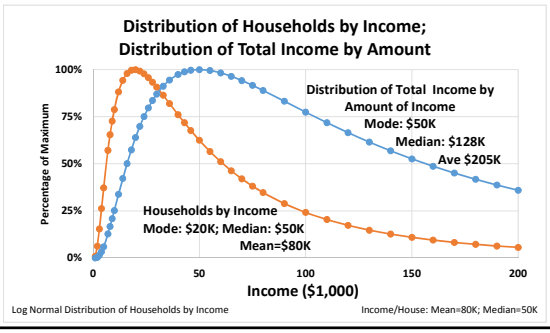
1C 2014 NNN2 9

## Distribution of Total Income



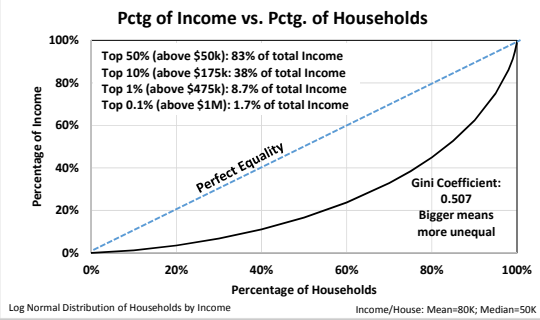

1C 2014 NNN2 10

## Distribution of Households and Total Income




1C 2014 NNN2 11

## Lorenz Curve and Gini Coefficient

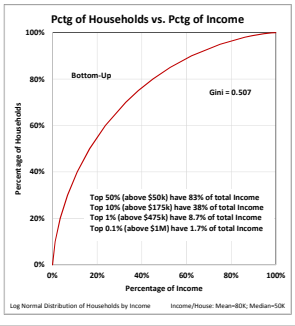



1C 2014 NNN2 12

## Champagne-Glass Distribution

The Gini coefficient is determined by the Mean#/Median# ratio.

The bigger this ratio the bigger the Gini coefficient and the greater the economic inequality.



1C 2014 NNN2 13

## As Mean-Median Ratio ↑ Rich get Richer (or vice-versa)

Log-normal distribution. Median HH income: $50K.

| | Top 5% | | Top 1% | | |
|---|---|---|---|---|---|
| Mean# | Min$ | %Income | Min$ | %Income | Gini |
| 55 | 103 | 11% | 138 | 2.9% | 0.24 |
| 60 | 135 | 15% | 204 | 4.2% | 0.33 |
| 65 | 165 | 18% | 270 | 5.5% | 0.39 |
| 70 | 193 | 20% | 337 | 6.6% | 0.44 |
| 75 | 220 | 23% | 406 | 7.7% | 0.48 |
| 80 | 246 | 25% | 477 | 8.7% | 0.51 |
| 85 | 272 | 27% | 549 | 9.7% | 0.53 |
| 90 | 298 | 29% | 623 | 10.7% | 0.56 |

1C 2014 NNN2 14

## As Mean-Median ratio rises, Modal Income may decrease!

| Median fixed at $50K | | | | Top 5% Households | | |
|---|---|---|---|---|---|---|
| Median | Ratio | Mean# | Mode# | Min$ | %Income | Gini |
| 50 | 1.2 | 60 | 35 | 135 | 15% | 0.33 |
| 50 | 1.3 | 65 | 30 | 165 | 18% | 0.39 |
| 50 | 1.4 | 70 | 26 | 193 | 20% | 0.44 |
| 50 | 1.5 | 75 | 22 | 220 | 23% | 0.48 |
| 50 | 1.6 | 80 | 20 | 246 | 25% | 0.51 |
| 50 | 1.7 | 85 | 17 | 272 | 27% | 0.53 |
| 50 | 1.8 | 90 | 15 | 298 | 29% | 0.56 |

Does this mean the poor get poorer as the rich get richer when median Income stays constant?

1C 2014 NNN2 15

## As Mean-Median ratio & Median ↑, Mode may increase

| | | | | ----- Top 5% ----- | | |
|---|---|---|---|---|---|---|
| Median | Ratio | Mean# | Mode# | Min$ | %Income | Gini |
| 40 | 1.2 | 48 | 28 | 108 | 15% | 0.33 |
| 50 | 1.3 | 65 | 30 | 165 | 18% | 0.39 |
| 60 | 1.4 | 84 | 31 | 231 | 20% | 0.44 |
| 70 | 1.5 | 105 | 31 | 308 | 23% | 0.48 |
| 80 | 1.6 | 128 | 31 | 394 | 25% | 0.51 |
| 90 | 1.7 | 153 | 31 | 490 | 27% | 0.53 |
| 100 | 1.8 | 180 | 31 | 595 | 29% | 0.56 |

What does this mean?

1C 2014 NNN2 16

## Share of Top 10%, Bottom 40% and their Palma Ratio

Palma ratio: [Share of top10%] / [Share of bottom 40%]. Cobham and Sumner (2014) argue that the Palma ratio is a more understandable measure of inequality than the Gini.

| | ---- Top 10% ---- | | -- Bottom 40% -- | | | |
|---|---|---|---|---|---|---|
| Mean# | Min$ | %Income | Max$ | %Income | Palma | Gini |
| 55 | 87 | 20% | 45 | 25% | 0.8 | 0.24 |
| 60 | 108 | 25% | 43 | 20% | 1.3 | 0.33 |
| 65 | 127 | 29% | 42 | 16% | 1.8 | 0.39 |
| 70 | 143 | 32% | 41 | 14% | 2.3 | 0.44 |
| 75 | 159 | 35% | 40 | 12% | 2.8 | 0.48 |
| 80 | 173 | 38% | 39 | 11% | 3.4 | 0.51 |
| 85 | 187 | 40% | 39 | 10% | 4.0 | 0.53 |
| | | Median Income: $50K | | | | |

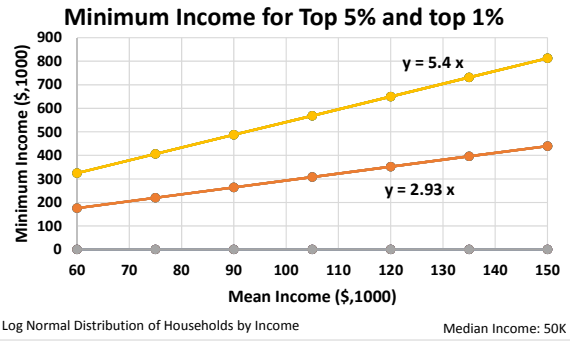
1C 2014 NNN2 17

## Share of Top 10%, Bottom 40% and their Palma Ratio

Palma and Gini are independent of the Median Income when the Mean-Median Income ratio is constant.

| | | | ---- Top 10% ---- | | -- Bottom 40% -- | | | |
|---|---|---|---|---|---|---|---|---|
| Median | Ratio | Mean# | Min$ | %Income | Max$ | %Income | Palma | Gini |
| 40 | 1.5 | 60 | 127 | 35% | 32 | 12% | 2.83 | 0.48 |
| 50 | 1.5 | 75 | 159 | 35% | 40 | 12% | 2.83 | 0.48 |
| 60 | 1.5 | 90 | 190 | 35% | 48 | 12% | 2.83 | 0.48 |
| 70 | 1.5 | 105 | 222 | 35% | 56 | 12% | 2.83 | 0.48 |
| 80 | 1.5 | 120 | 254 | 35% | 64 | 12% | 2.83 | 0.48 |
| 90 | 1.5 | 135 | 285 | 35% | 72 | 12% | 2.83 | 0.48 |
| 100 | 1.5 | 150 | 317 | 35% | 80 | 12% | 2.83 | 0.48 |
| Constant Mean-Median Ratio | | | | | | | | |

1C 2014 NNN2 18




Log Normal Distribution of Households by Income

Median Income: 50K

1C 2014 NNN2 19

## Which parameters best model US household incomes?

US Median Income (Table 691*)
- $46,089 in 1970; $50,303 in 2008

Share of Total Income by Top 5% (Table 693*)
- 16.6% in 1970; 21.5% in 2008

Best log-normal fits:
- 1970 Median 46K, Mean 53K: Ratio = 1.15
- 2008 *Median 50K, Mean 73K*; Ratio = 1.46

* 2011 US Statistical Abstract (2008 dollars).

1C 2014 NNN2 20

## Distinguish whole & part

Consider a lognormal distribution of family incomes with a median of $50K and a mean of $80K.

What percentage
- of income is held by the top 5% of families?
- of families hold the top 5% of income?

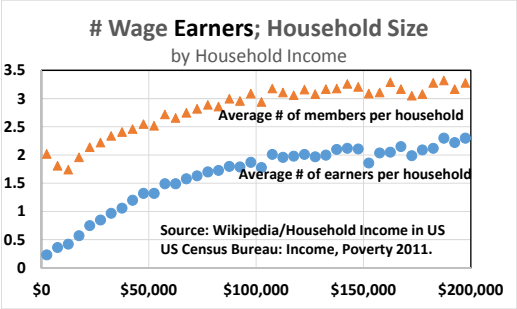
Is there a difference in these percentages? Why?

Which one is generally larger? Why?

What are some other causes of income differences?

1C 2014 NNN2 21

## Explore the Causes of Income Differences




1C 2014 NNN2 22

## Explore the Causes of Income Differences

| Type of Household | Lowest fifth | Second fifth | Middle fifth | Fourth fifth | Highest fifth | Top 5% |
|---|---|---|---|---|---|---|
| Married couple families | 17% | 36% | 48% | 65% | 78% | 82% |
| Single-male family | 4% | 6% | 6% | 5% | 4% | 2% |
| Single-female family | 20% | 17% | 14% | 9% | 5% | 4% |
| Non-family households | 60% | 42% | 32% | 21% | 13% | 12% |
| TOTAL | 100% | 100% | 100% | 100% | 100% | 100% |
| Mean # of income earners | 0.4 | 0.9 | 1.3 | 1.7 | 2 | 2.1 |

1C 2014 NNN2 23

## Conclusion

Using the LogNormal distributions provides a principled way students can explore a plausible distribution of incomes.

Allows students to explore the difference between part and whole when using percentage grammar.

1C 2014 NNN2 24

## Bibliography

Aitchison J and JAC Brown (1957). *The Log-normal Distribution*. Cambridge (UK): Cambridge University Press. Searchable copy at Google Books: http://books.google.com/books?id=Kus8AAAAIAAJ

Cobham, Alex and Andy Sumner (2014). Is inequality all about the tails?: The Palma measure of income inequality. Significance. Volume 11 Issue 1. www.significancemagazine.org/details/magazine/5871201/Is-inequality-all-about-the-tails-The-Palma-measure-of-income-inequality.html

Limpert, E., W.A. Stahel and M. Abbt (2001). Log-normal Distributions across the Sciences: Keys and Clues. *Bioscience* 51, No 5, May 2001, 342-352. Copy at http://stat.ethz.ch/~stahel/lognormal/bioscience.pdf

Schield, Milo (2013) Creating a Log-Normal Distribution using Excel 2013. www.statlit.org/pdf/Create-LogNormal-Excel2013-Demo-6up.pdf

Stahel, Werner (2014). Website: http://stat.ethz.ch/~stahel

Univ. Denver (2014). Using the LogNormal Distribution. Copy at http://www.du.edu/ifs/help/understand/economy/poverty/lognormal.html

Wikipedia. LogNormal Distribution.

# Exploring Lognormal Incomes

**Milo Schield**
**Augsburg College**
**Editor: www.StatLit.org**
**US Rep: International Statistical Literacy Project**

**11 October 2014**

**National Numeracy Network**

**www.StatLit.org/pdf/2014-Schield-NNN2-Slides.pdf**

**www.StatLit.org/Excel/Create-LogNormal-Incomes-Excel2013.xlsx**

# Log-Normal Distributions

A Log-Normal distribution is generated from a normal with mu = Ln(Median) and sigma = Sqrt[2*Ln(Mean/Median)].

The lognormal is always positive and right-skewed.

Examples:

- Incomes (bottom 97%), assets, size of cities
- Weight and blood pressure of humans (by gender)

Benefit:

- calculate the share of total income held by the top X%
- calculate Gini Coefficient,
- explore effects of change in mean-median ratio.

# Log-Normal Distributions

"In many ways, it [the Log-Normal] has remained the Cinderella of distributions, the interest of writers in the learned journals being curiously sporadic and that of the authors of statistical test-books but faintly aroused."

"We … state our belief that the lognormal is as fundamental a distribution in statistics as is the normal, despite the stigma of the derivative nature of its name."

Aitchison and Brown (1957). P 1.

# Lognormal and Excel

Use Excel to focus on the model and the results.

Excel has two Log-Normal functions:

- Standard: =LOGNORM.DIST(X, mu, sigma, k)
  k=0 for PDF; k=1 for CDF.
- Inverse: =LOGNORM.INV(X, mu, sigma)

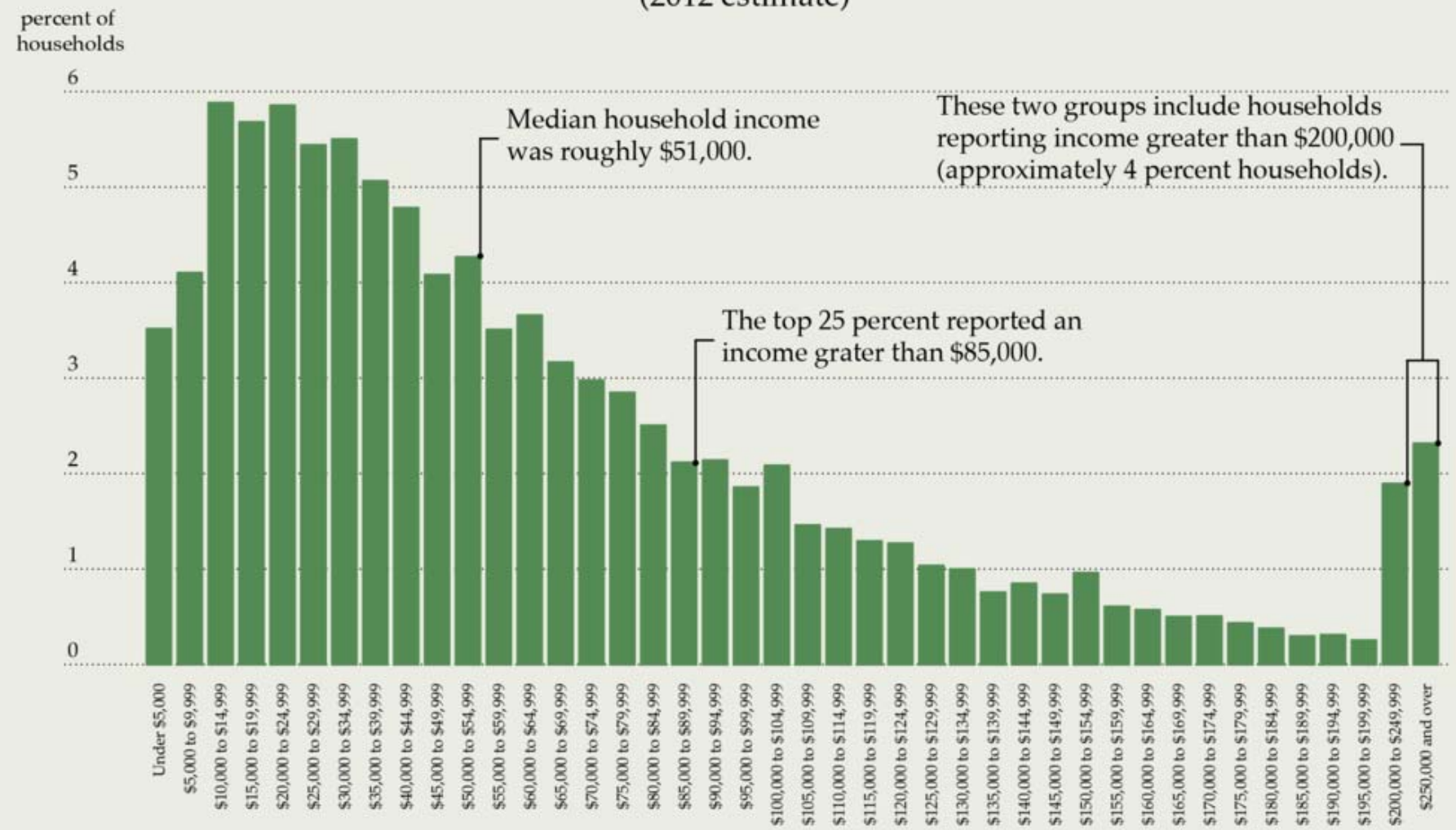
Use Standard to calculate/graph the PDF and CDF.

Use Inverse to find cutoffs: quartiles, to 1%, etc.

Use Excel to create graphs that show comparisons.

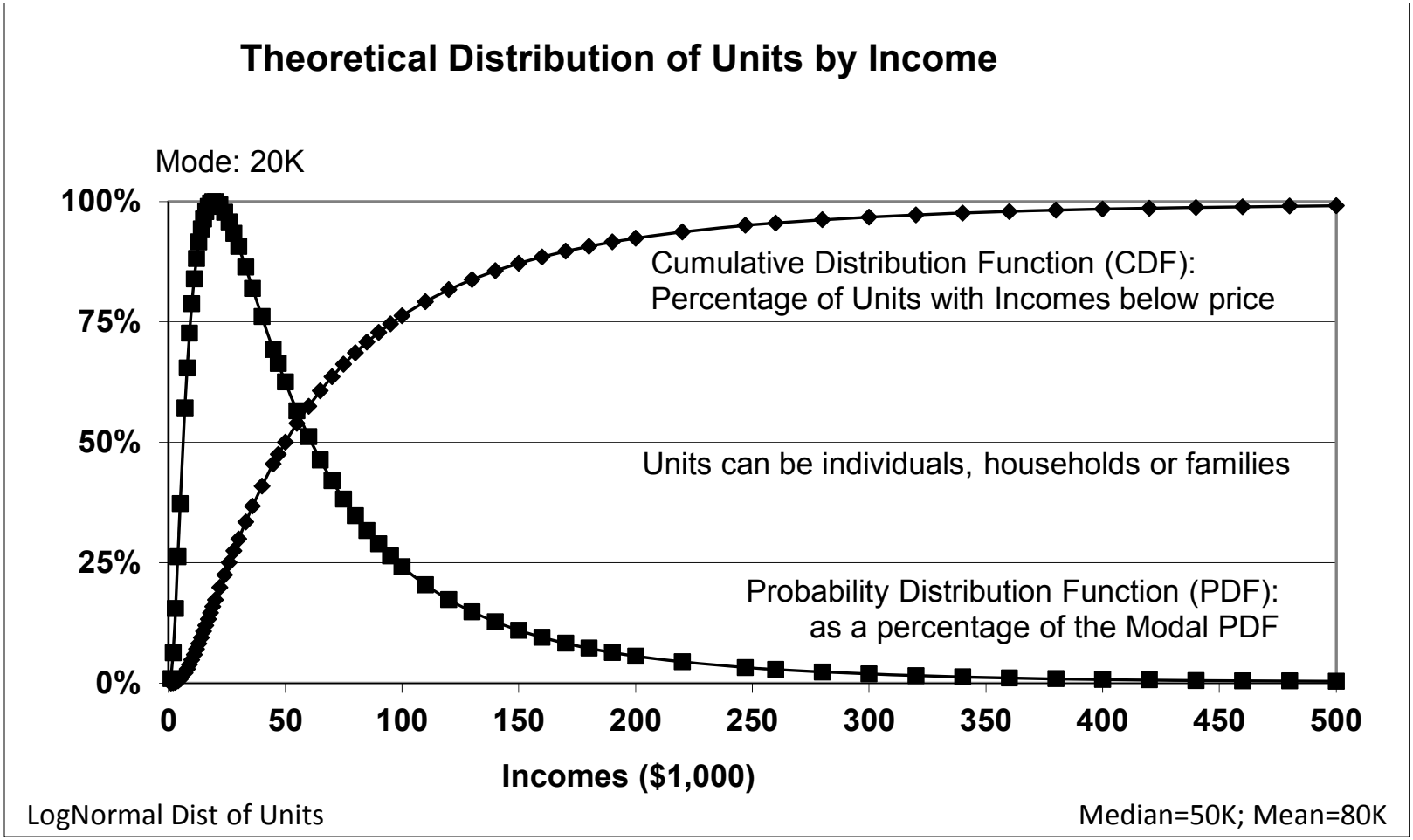
# Distribution of annual household income in the United States

(2012 estimate)




Source: U.S. Census Bureau, Current Population Survey, 2012 Annual Social and Economic Supplement

# Log-Normal Distribution of Units



# Paired Distributions

For anything that is distributed by X, there are always two distributions:

1. Distribution of subjects by X
2. Distribution of total X by X.

Sometime we ignore the 2nd: height or weight.

Sometimes we care about the 2nd: income or assets.

Surprise: If the 1st is lognormal, so is the 2nd.

# Distribution of Households and Total Income by Income

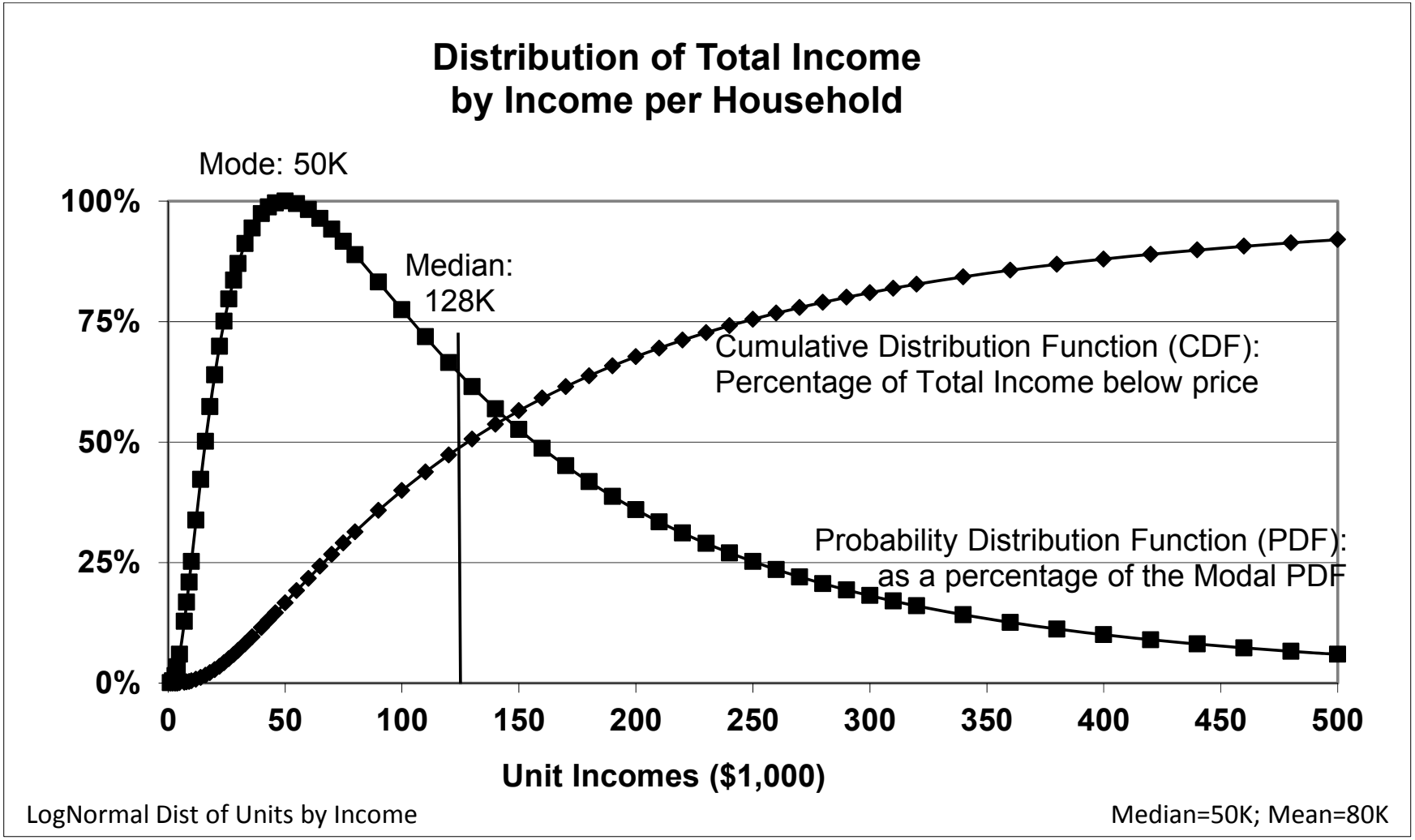
Suppose the distribution of households by income is log-normal with normal parameters mu# and sigma#.

Then the distribution of total income by amount has a log-normal distribution with these parameters:
mu$ = mu# + sigma#^2;  sigma$ = sigma#.

See Aitchison and Brown (1963) p. 158.

Special thanks to Mohammod Irfan (Denver University) for his help on this topic.

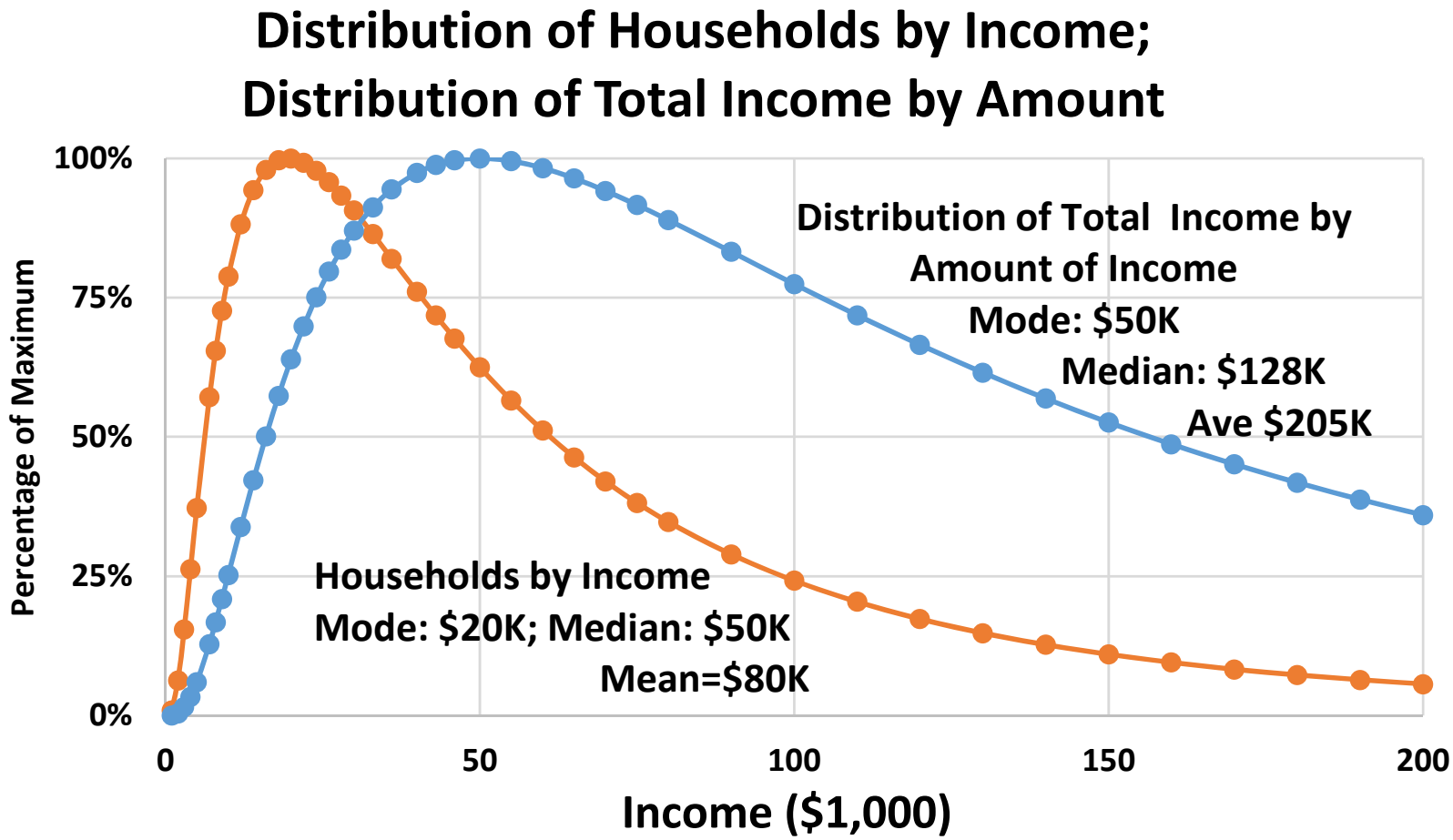
# Distribution of Total Income




LogNormal Dist of Units by Income

Median=50K; Mean=80K

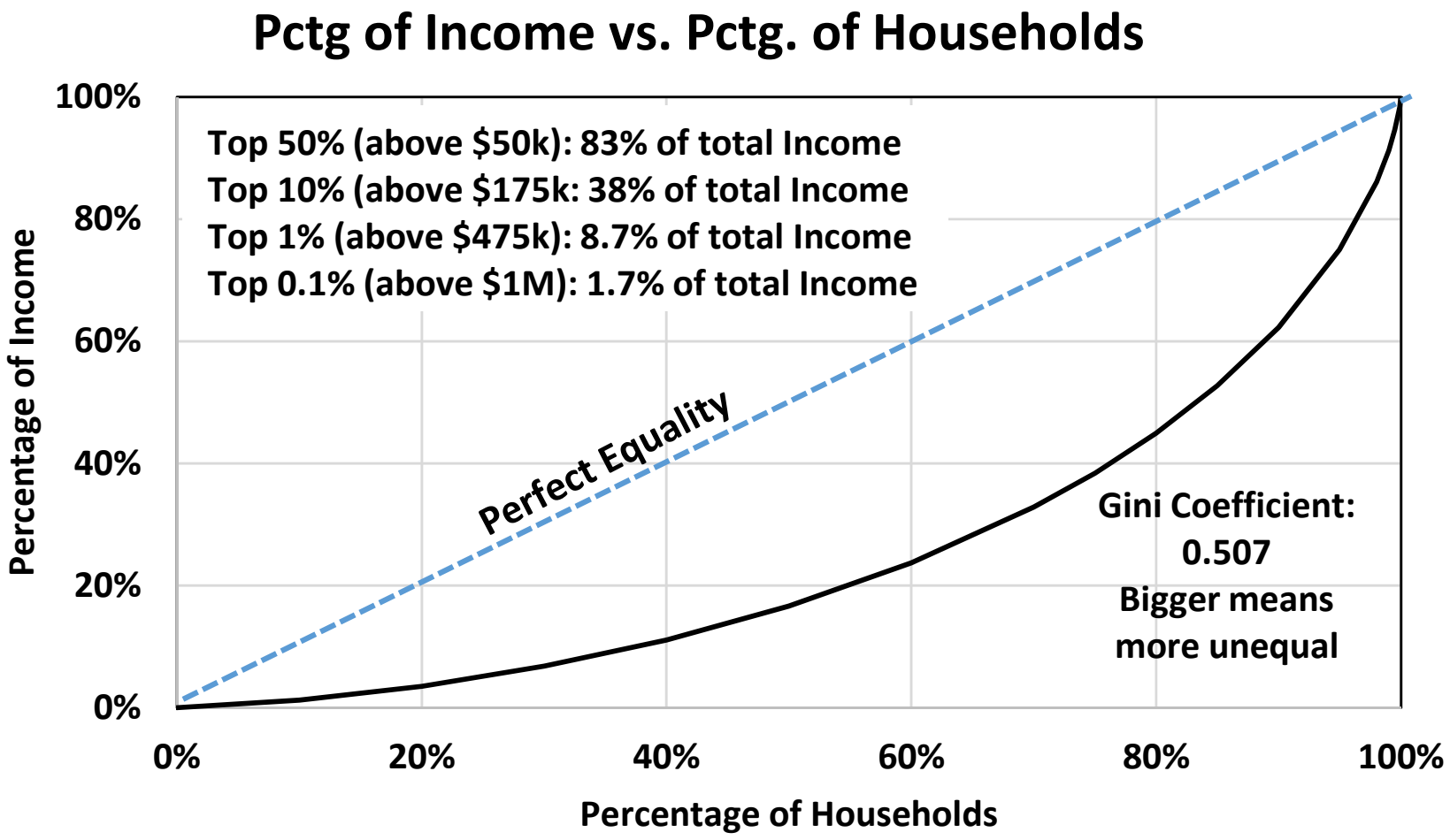
# Distribution of Households and Total Income




Log Normal Distribution of Households by Income

Income/House: Mean=80K; Median=50K

# Lorenz Curve and Gini Coefficient



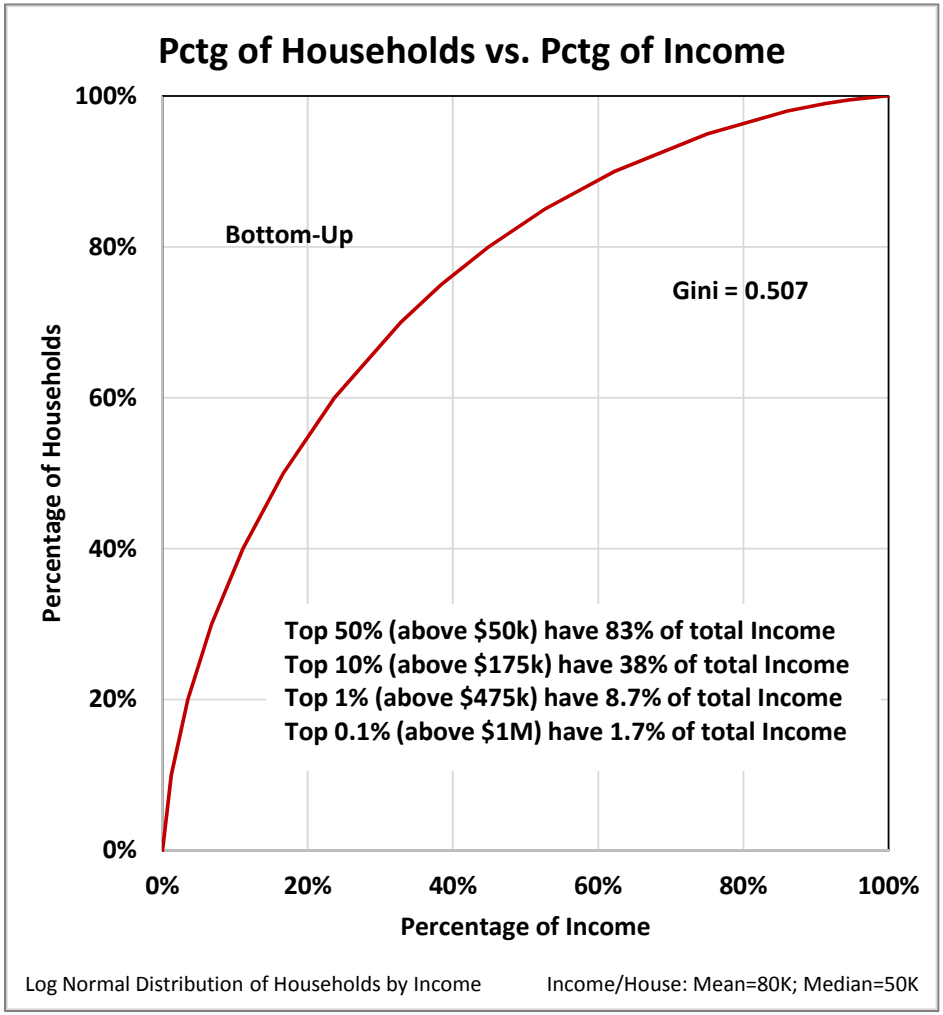

Log Normal Distribution of Households by Income

Income/House: Mean=80K; Median=50K

# Champagne-Glass Distribution

The Gini coefficient is determined by the Mean#/Median# ratio.

The bigger this ratio the bigger the Gini coefficient and the greater the economic inequality.



# As Mean-Median Ratio ↑ Rich get Richer (or vice-versa)

Log-normal distribution. Median HH income: $50K.

| | Top 5% | | Top 1% | | |
|---|---|---|---|---|---|
| Mean# | Min$ | %Income | Min$ | %Income | Gini |
| 55 | 103 | 11% | 138 | 2.9% | 0.24 |
| 60 | 135 | 15% | 204 | 4.2% | 0.33 |
| 65 | 165 | 18% | 270 | 5.5% | 0.39 |
| 70 | 193 | 20% | 337 | 6.6% | 0.44 |
| 75 | 220 | 23% | 406 | 7.7% | 0.48 |
| 80 | 246 | 25% | 477 | 8.7% | 0.51 |
| 85 | 272 | 27% | 549 | 9.7% | 0.53 |
| 90 | 298 | 29% | 623 | 10.7% | 0.56 |

# As Mean-Median ratio rises, Modal Income may decrease!

| Median fixed at $50K | | | | Top 5% Households | | |
|---|---|---|---|---|---|---|
| Median | Ratio | Mean# | Mode# | Min$ | %Income | Gini |
| 50 | 1.2 | 60 | 35 | 135 | 15% | 0.33 |
| 50 | 1.3 | 65 | 30 | 165 | 18% | 0.39 |
| 50 | 1.4 | 70 | 26 | 193 | 20% | 0.44 |
| 50 | 1.5 | 75 | 22 | 220 | 23% | 0.48 |
| 50 | 1.6 | 80 | 20 | 246 | 25% | 0.51 |
| 50 | 1.7 | 85 | 17 | 272 | 27% | 0.53 |
| 50 | 1.8 | 90 | 15 | 298 | 29% | 0.56 |

Does this mean the poor get poorer as the rich get richer when median Income stays constant?

# As Mean-Median ratio & Median ↑, Mode may increase

| | | | | ----- Top 5% ----- | | |
|---|---|---|---|---|---|---|
| Median | Ratio | Mean# | Mode# | Min$ | %Income | Gini |
| 40 | 1.2 | 48 | 28 | 108 | 15% | 0.33 |
| 50 | 1.3 | 65 | 30 | 165 | 18% | 0.39 |
| 60 | 1.4 | 84 | 31 | 231 | 20% | 0.44 |
| 70 | 1.5 | 105 | 31 | 308 | 23% | 0.48 |
| 80 | 1.6 | 128 | 31 | 394 | 25% | 0.51 |
| 90 | 1.7 | 153 | 31 | 490 | 27% | 0.53 |
| 100 | 1.8 | 180 | 31 | 595 | 29% | 0.56 |

What does this mean?

# Share of Top 10%, Bottom 40% and their Palma Ratio

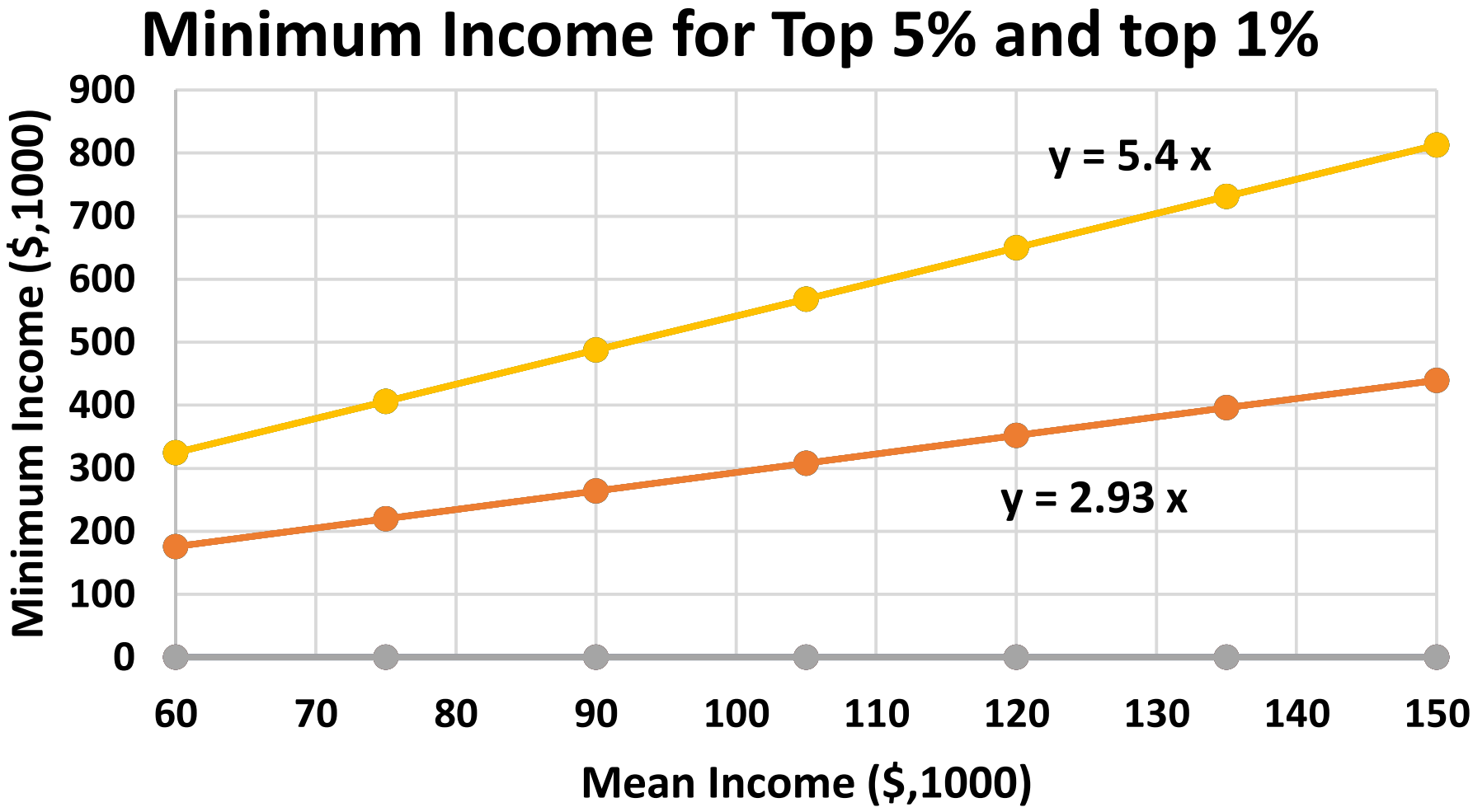
Palma ratio: [Share of top10%] / [Share of bottom 40%]. Cobham and Sumner (2014) argue that the Palma ratio is a more understandable measure of inequality than the Gini.

| | ---- Top 10% ---- | | -- Bottom 40% -- | | | |
|---|---|---|---|---|---|---|
| Mean# | Min$ | %Income | Max$ | %Income | Palma | Gini |
| 55 | 87 | 20% | 45 | 25% | 0.8 | 0.24 |
| 60 | 108 | 25% | 43 | 20% | 1.3 | 0.33 |
| 65 | 127 | 29% | 42 | 16% | 1.8 | 0.39 |
| 70 | 143 | 32% | 41 | 14% | 2.3 | 0.44 |
| 75 | 159 | 35% | 40 | 12% | 2.8 | 0.48 |
| 80 | 173 | 38% | 39 | 11% | 3.4 | 0.51 |
| 85 | 187 | 40% | 39 | 10% | 4.0 | 0.53 |
| | | **Median Income: $50K** | | | | |

# Share of Top 10%, Bottom 40% and their Palma Ratio

Palma and Gini are independent of the Median Income when the Mean-Median Income ratio is constant.

| | | | ---- Top 10% ---- | | -- Bottom 40% -- | | | |
|---|---|---|---|---|---|---|---|---|
| Median | Ratio | Mean# | Min$ | %Income | Max$ | %Income | Palma | Gini |
| 40 | 1.5 | 60 | 127 | 35% | 32 | 12% | 2.83 | 0.48 |
| 50 | 1.5 | 75 | 159 | 35% | 40 | 12% | 2.83 | 0.48 |
| 60 | 1.5 | 90 | 190 | 35% | 48 | 12% | 2.83 | 0.48 |
| 70 | 1.5 | 105 | 222 | 35% | 56 | 12% | 2.83 | 0.48 |
| 80 | 1.5 | 120 | 254 | 35% | 64 | 12% | 2.83 | 0.48 |
| 90 | 1.5 | 135 | 285 | 35% | 72 | 12% | 2.83 | 0.48 |
| 100 | 1.5 | 150 | 317 | 35% | 80 | 12% | 2.83 | 0.48 |
| **Constant Mean-Median Ratio** | | | | | | | | |




Log Normal Distribution of Households by Income

Median Income: 50K

# Which parameters best model US household incomes?

US Median Income (Table 691*)

- $46,089 in 1970; $50,303 in 2008

Share of Total Income by Top 5% (Table 693*)

- 16.6% in 1970; 21.5% in 2008

Best log-normal fits:

- 1970 Median 46K, Mean 53K: Ratio = 1.15
- 2008 *Median 50K, Mean 73K*; Ratio = 1.46

* 2011 US Statistical Abstract (2008 dollars).

# Distinguish whole & part

Consider a lognormal distribution of family incomes with a median of $50K and a mean of $80K.

What percentage

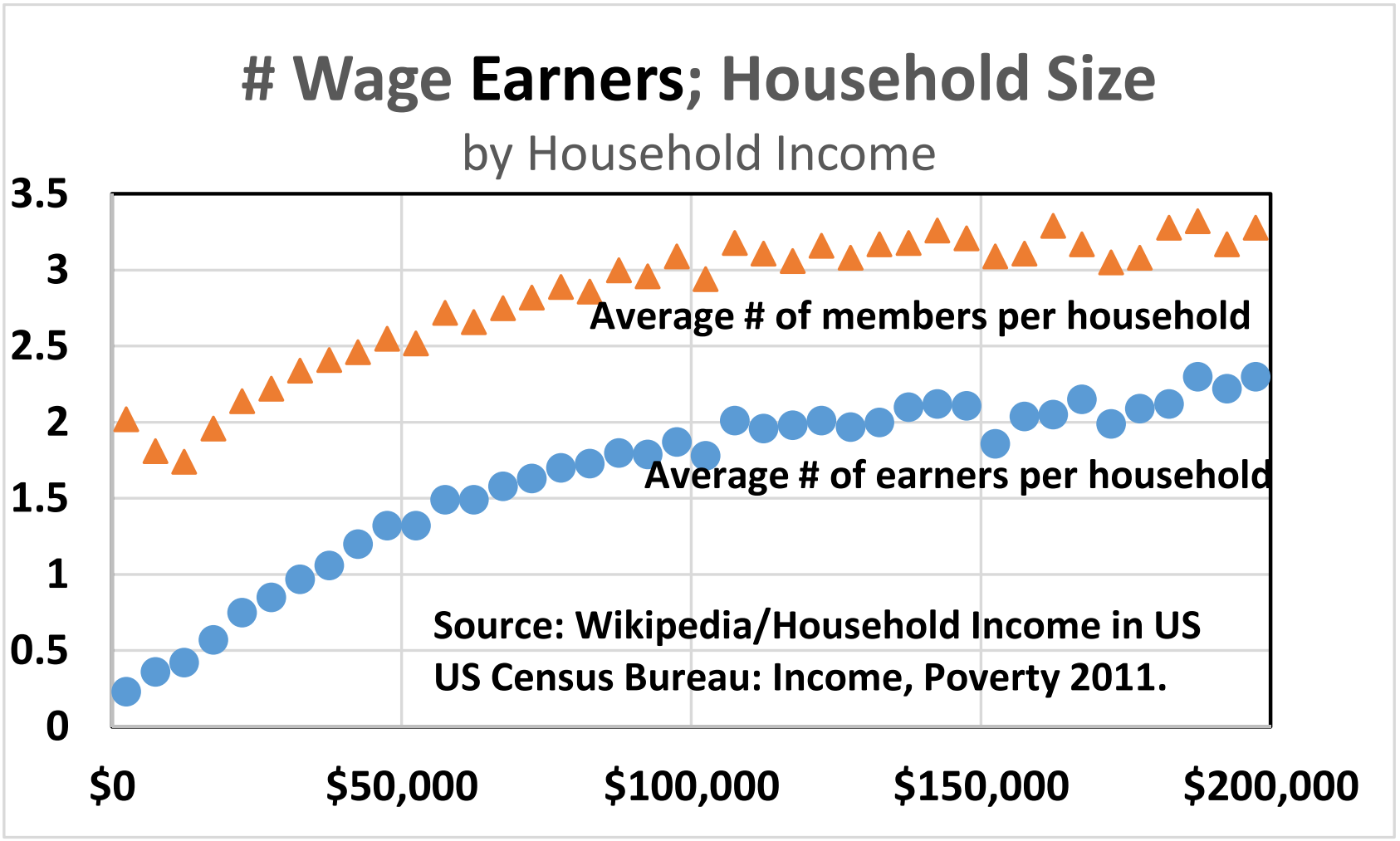
- of income is held by the top 5% of families?
- of families hold the top 5% of income?

Is there a difference in these percentages? Why?

Which one is generally larger? Why?

What are some other causes of income differences?

# Explore the Causes of Income Differences



# Explore the Causes of Income Differences

| Type of Household | Lowest fifth | Second fifth | Middle fifth | Fourth fifth | Highest fifth | Top 5% |
|---|---|---|---|---|---|---|
| Married couple families | 17% | 36% | 48% | 65% | 78% | 82% |
| Single-male family | 4% | 6% | 6% | 5% | 4% | 2% |
| Single-female family | 20% | 17% | 14% | 9% | 5% | 4% |
| Non-family households | 60% | 42% | 32% | 21% | 13% | 12% |
| TOTAL | 100% | 100% | 100% | 100% | 100% | 100% |
| Mean # of income earners | 0.4 | 0.9 | 1.3 | 1.7 | 2 | 2.1 |

# Conclusion

Using the LogNormal distributions provides a principled way students can explore a plausible distribution of incomes.

Allows students to explore the difference between part and whole when using percentage grammar.

# Bibliography

Aitchison J and JAC Brown (1957). *The Log-normal Distribution*. Cambridge (UK): Cambridge University Press. Searchable copy at Google Books: http://books.google.com/books?id=Kus8AAAAIAAJ

Cobham, Alex and Andy Sumner (2014). Is inequality all about the tails?: The Palma measure of income inequality. Significance. Volume 11 Issue 1. www.significancemagazine.org/details/magazine/5871201/Is-inequality-all-about-the-tails-The-Palma-measure-of-income-inequality.html

Limpert, E., W.A. Stahel and M. Abbt (2001). Log-normal Distributions across the Sciences: Keys and Clues. *Bioscience* 51, No 5, May 2001, 342-352. Copy at http://stat.ethz.ch/~stahel/lognormal/bioscience.pdf

Schield, Milo (2013) Creating a Log-Normal Distribution using Excel 2013. www.statlit.org/pdf/Create-LogNormal-Excel2013-Demo-6up.pdf

Stahel, Werner (2014). Website: http://stat.ethz.ch/~stahel

Univ. Denver (2014). Using the LogNormal Distribution. Copy at http://www.du.edu/ifs/help/understand/economy/poverty/lognormal.html

Wikipedia. LogNormal Distribution.