Fish Consumption Rates

Two sampling issues examined

Don A. Essig, Idaho DEQ

Two questions I want to examine with you

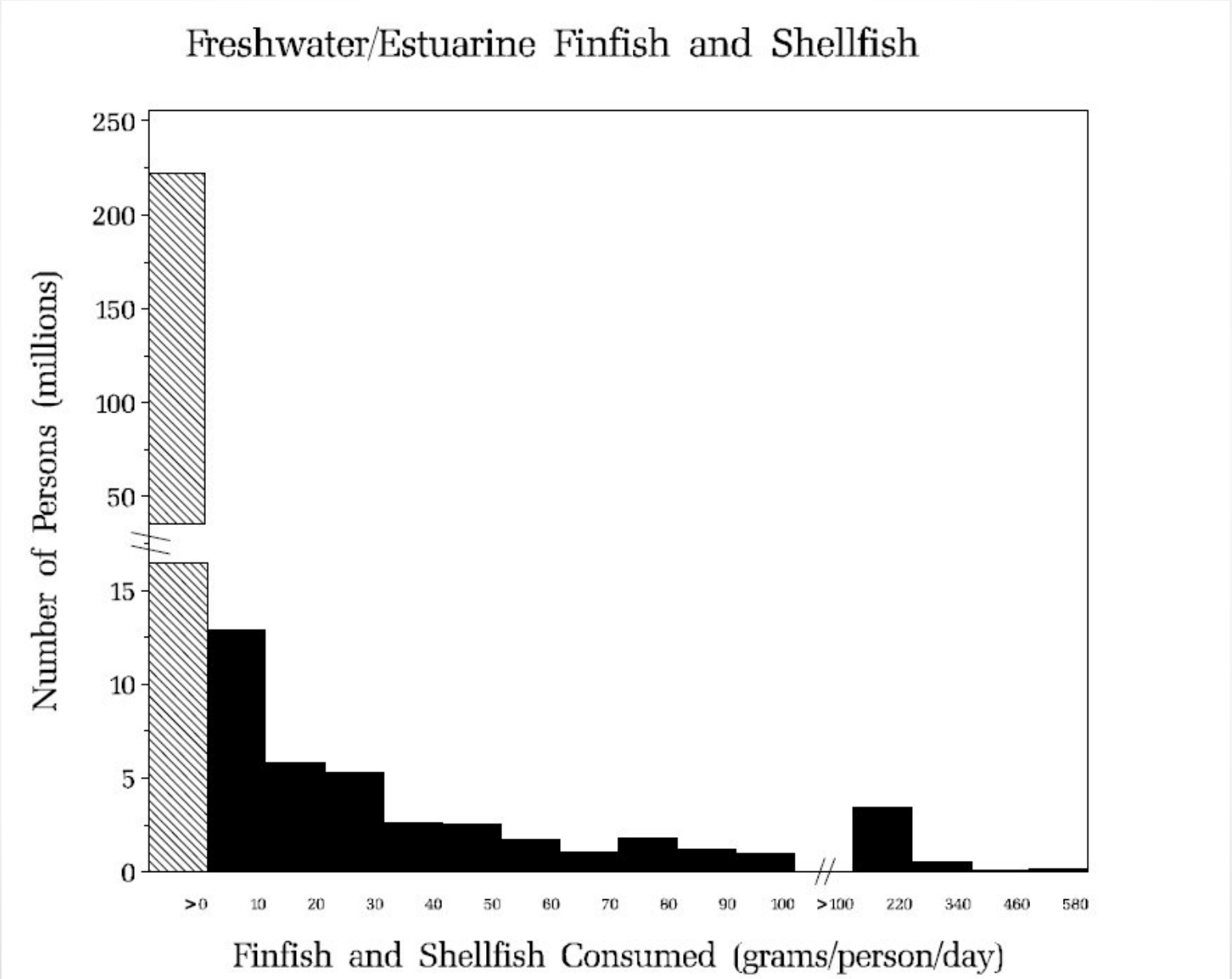
1. What happens if we mistakenly identify some consumers of fish as non-consumers of fish?

2. What happens if we miss some hard-to-reach people, AND they are high-end consumers?

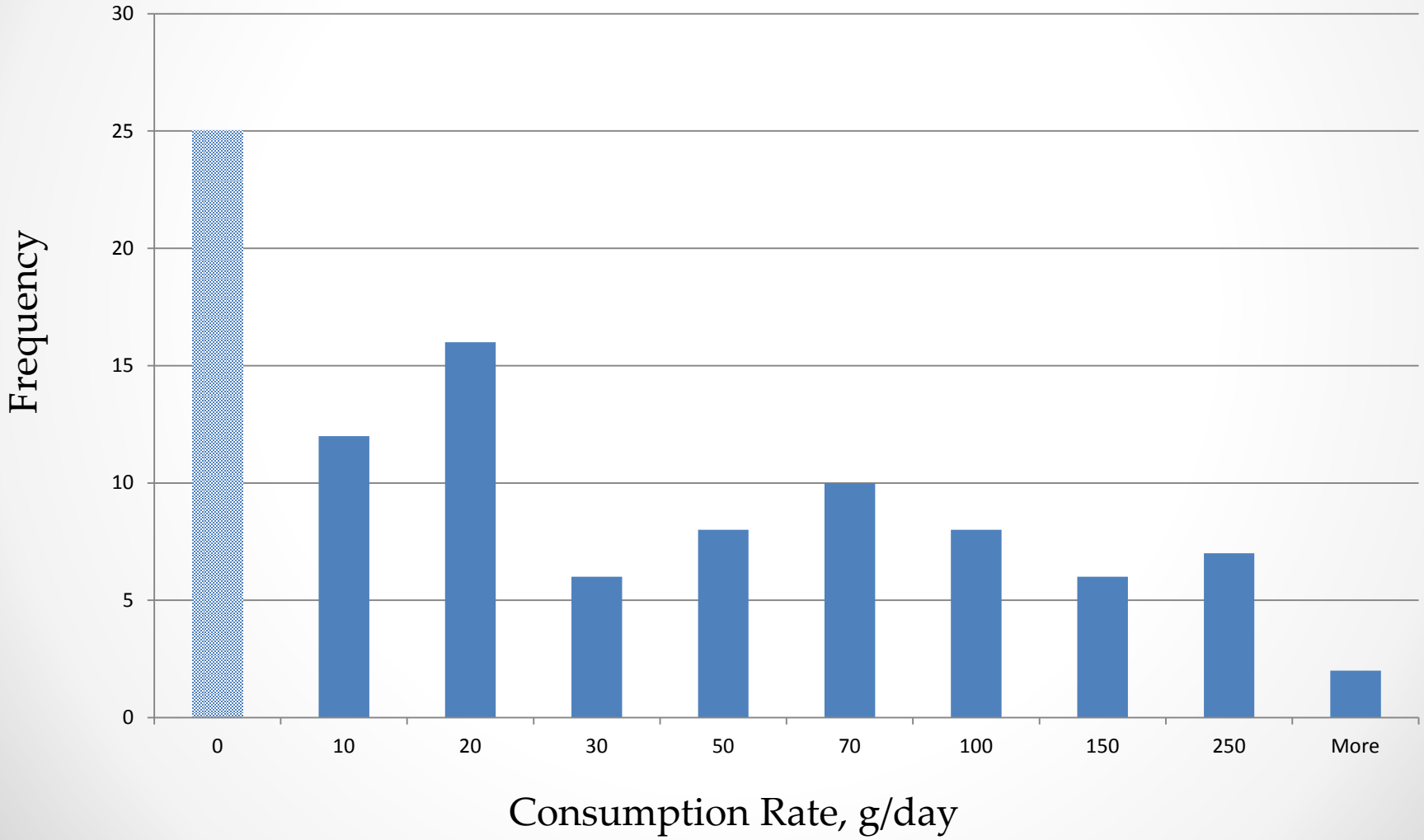
We will look at & compare 3 hypothetical distributions

1. A hypothetical base distribution, with 25% non-consumers
2. A modified base distribution in which 3/5ths of the apparent non-consumers actually consume fish, albeit at a low rate
3. A modified base distribution in which 5% more people are surveyed, whose consumption rates are in the upper half of the base distribution

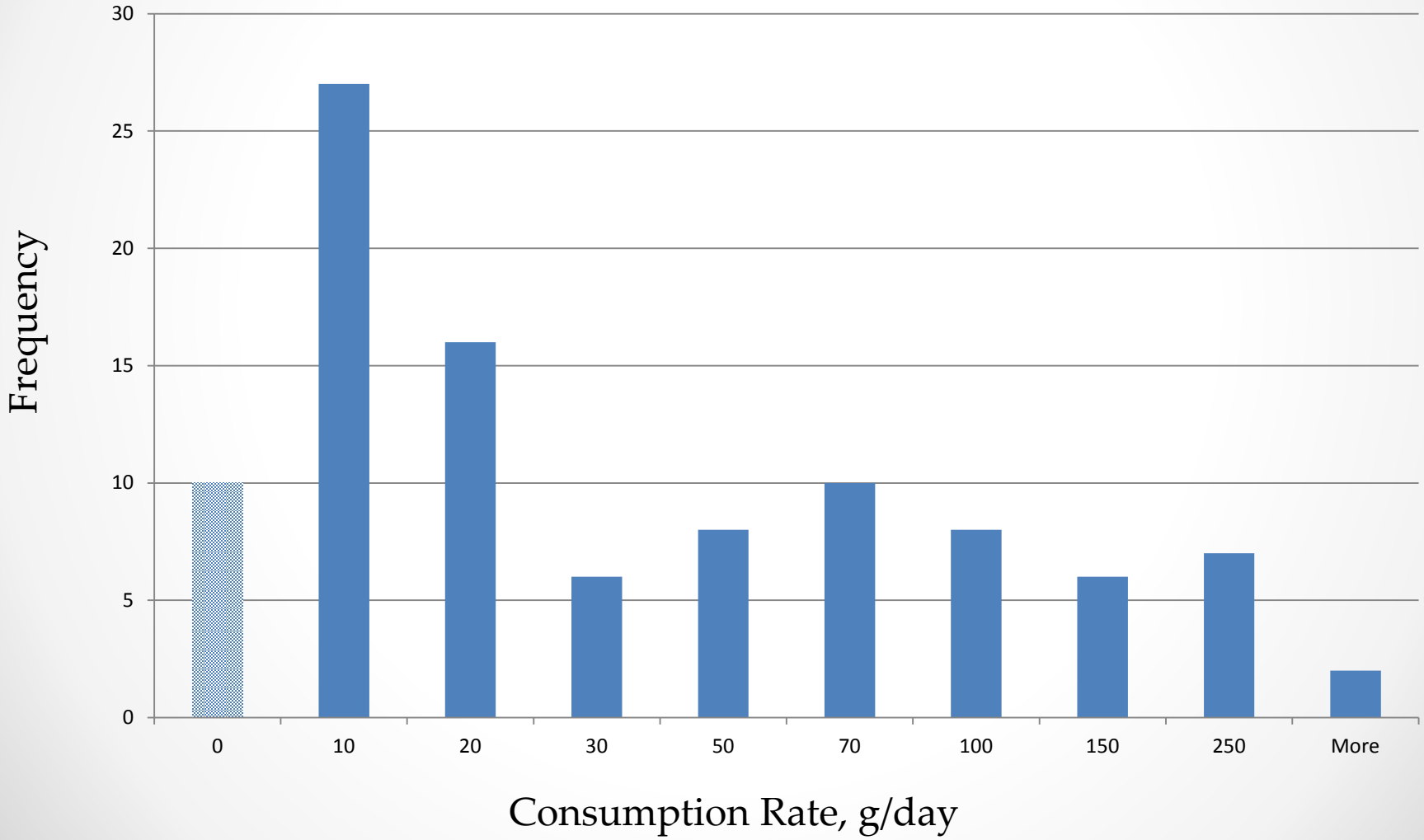
Figure 5-1-1-1 From EPA 2002



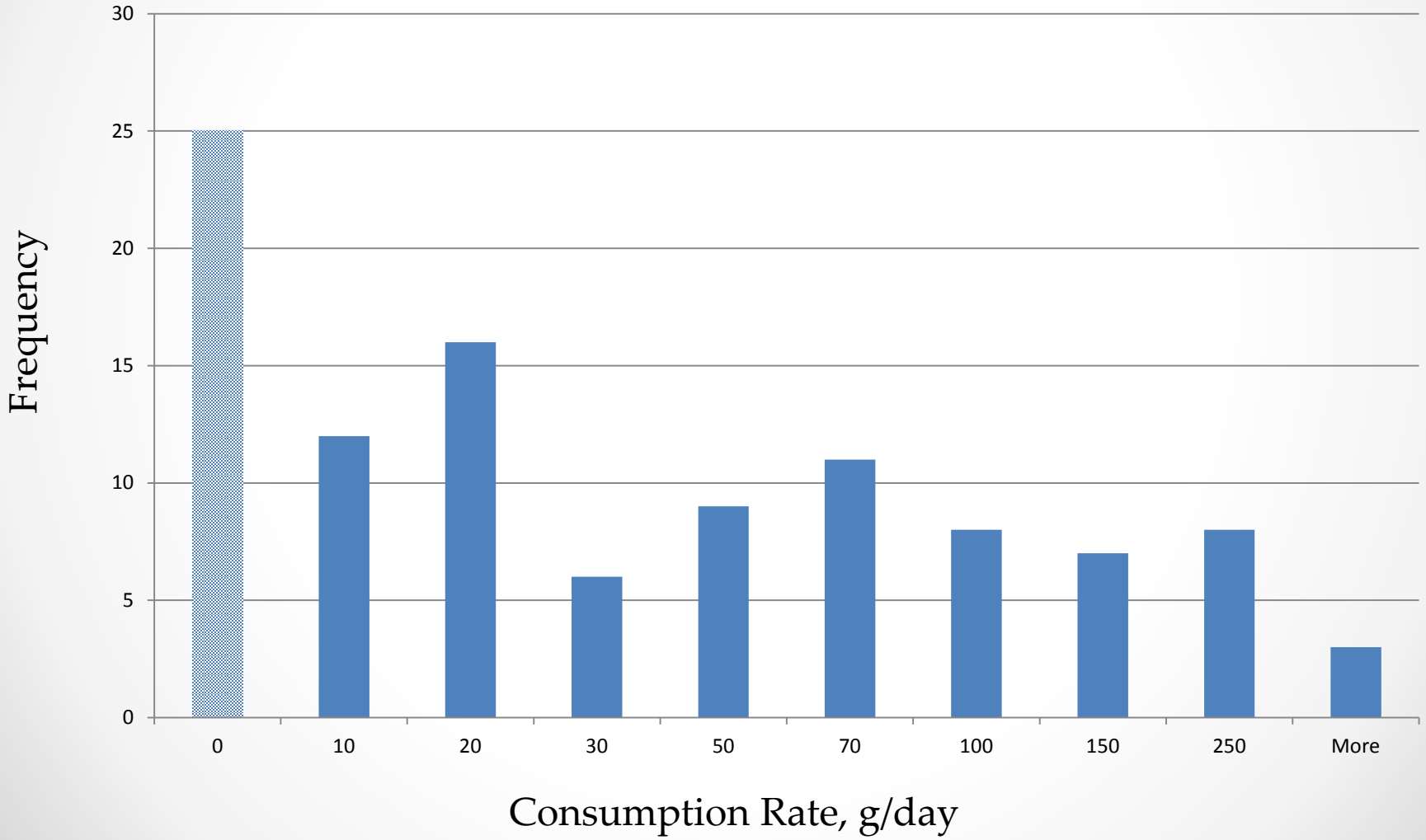
Hypothetical fish consumption distribution #1



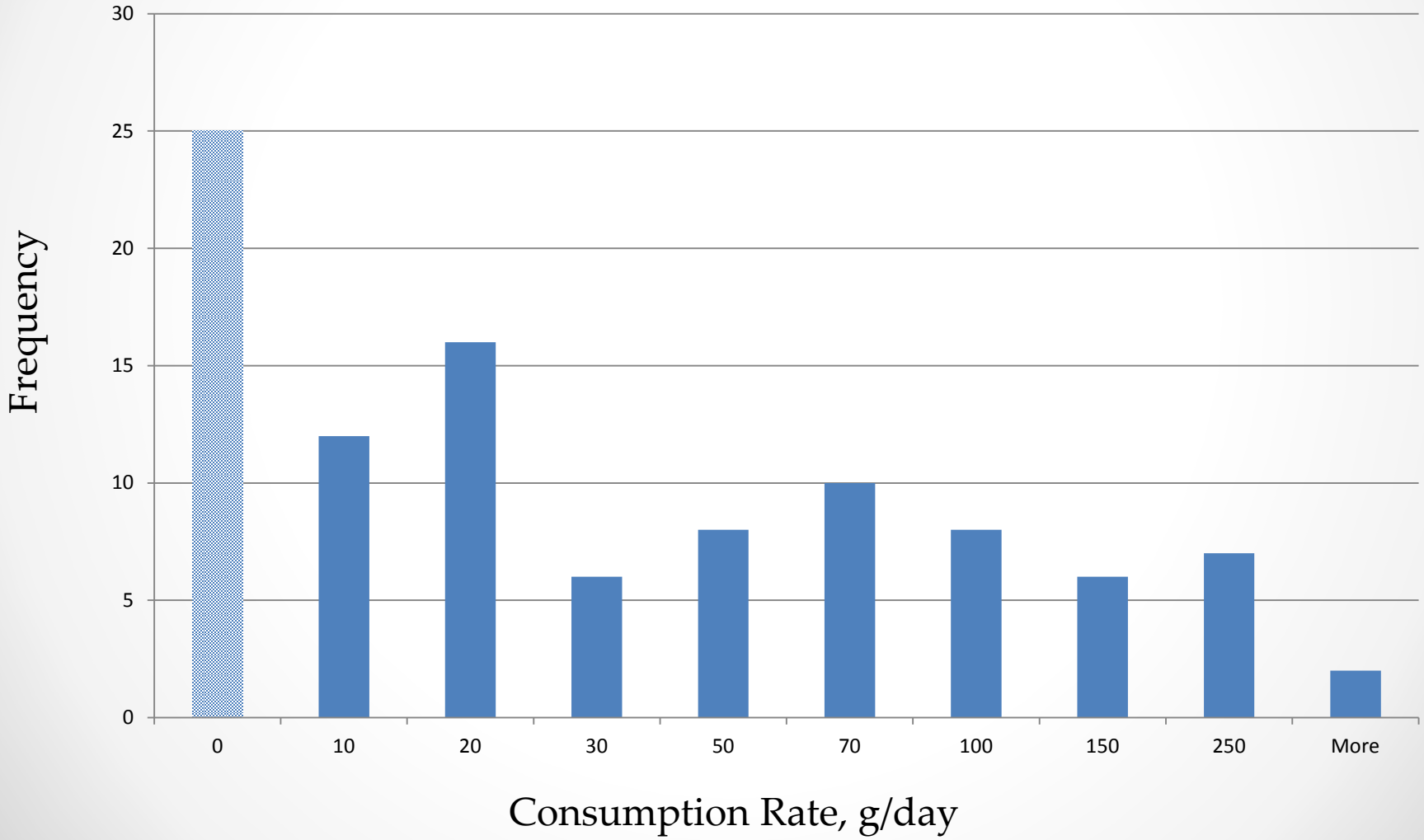
Hypothetical fish consumption distribution #2



Hypothetical fish consumption distribution #3



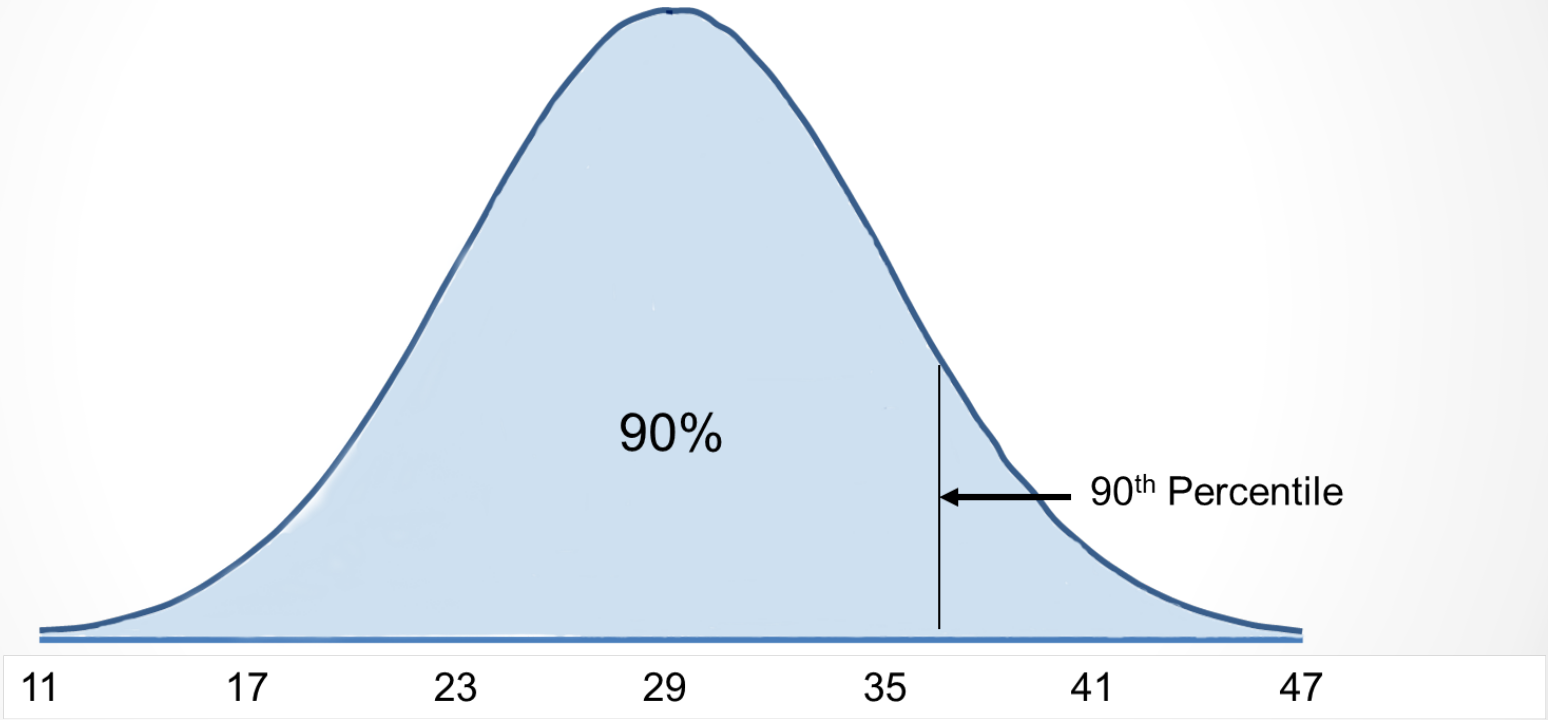
Hypothetical fish consumption distribution #1



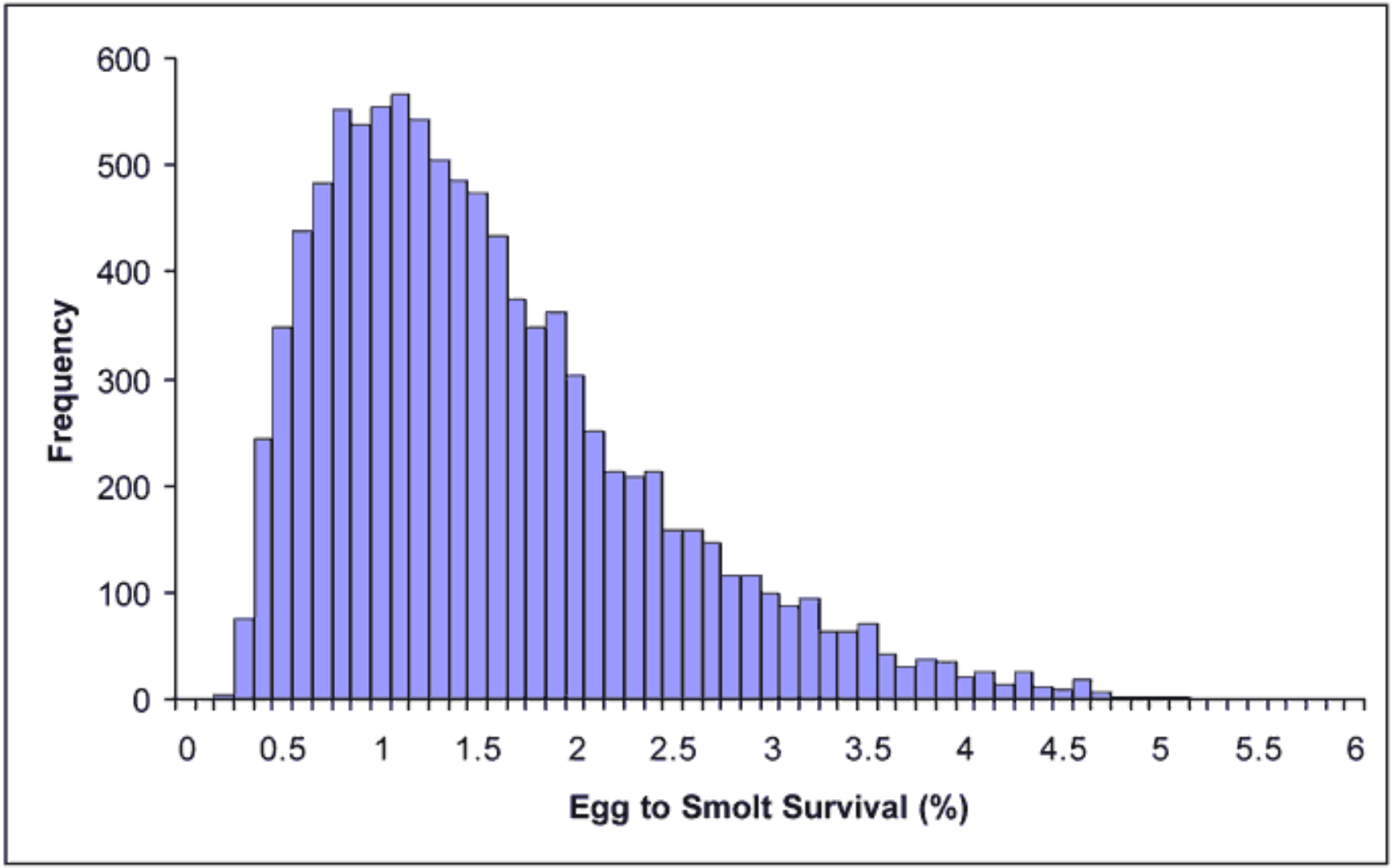
What Statistics?

- For fish consumption rates that go into risk assessment or water quality criteria we generally focus on the upper half of the distribution.
- But there are many statistics that can be used to describe the data
- We'll look at these five:
  1. Median, or 50th percentile
  2. Mean
  3. 90th percentile
  4. 95th percentile
  5. 99th percentile

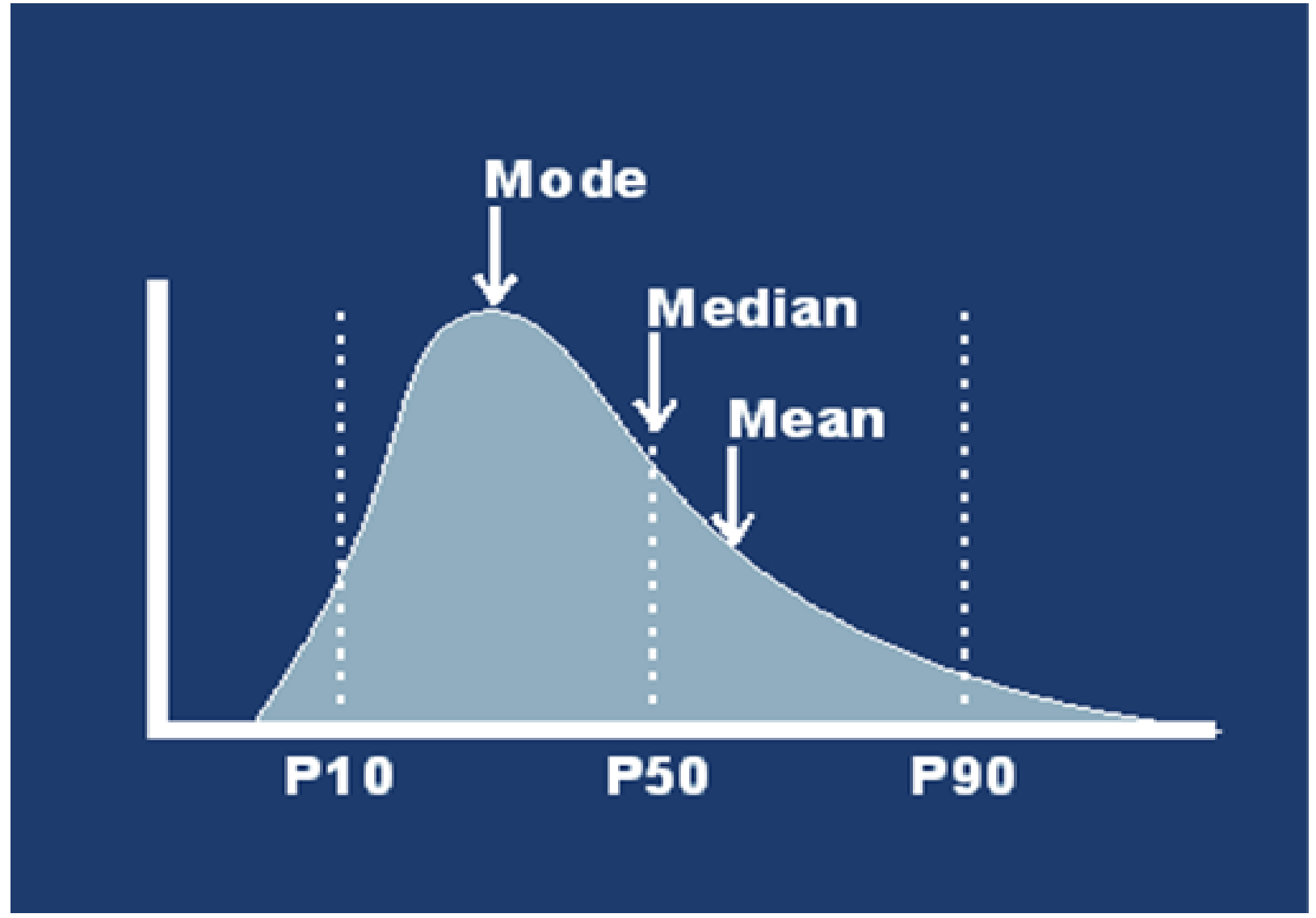
Normal Distribution

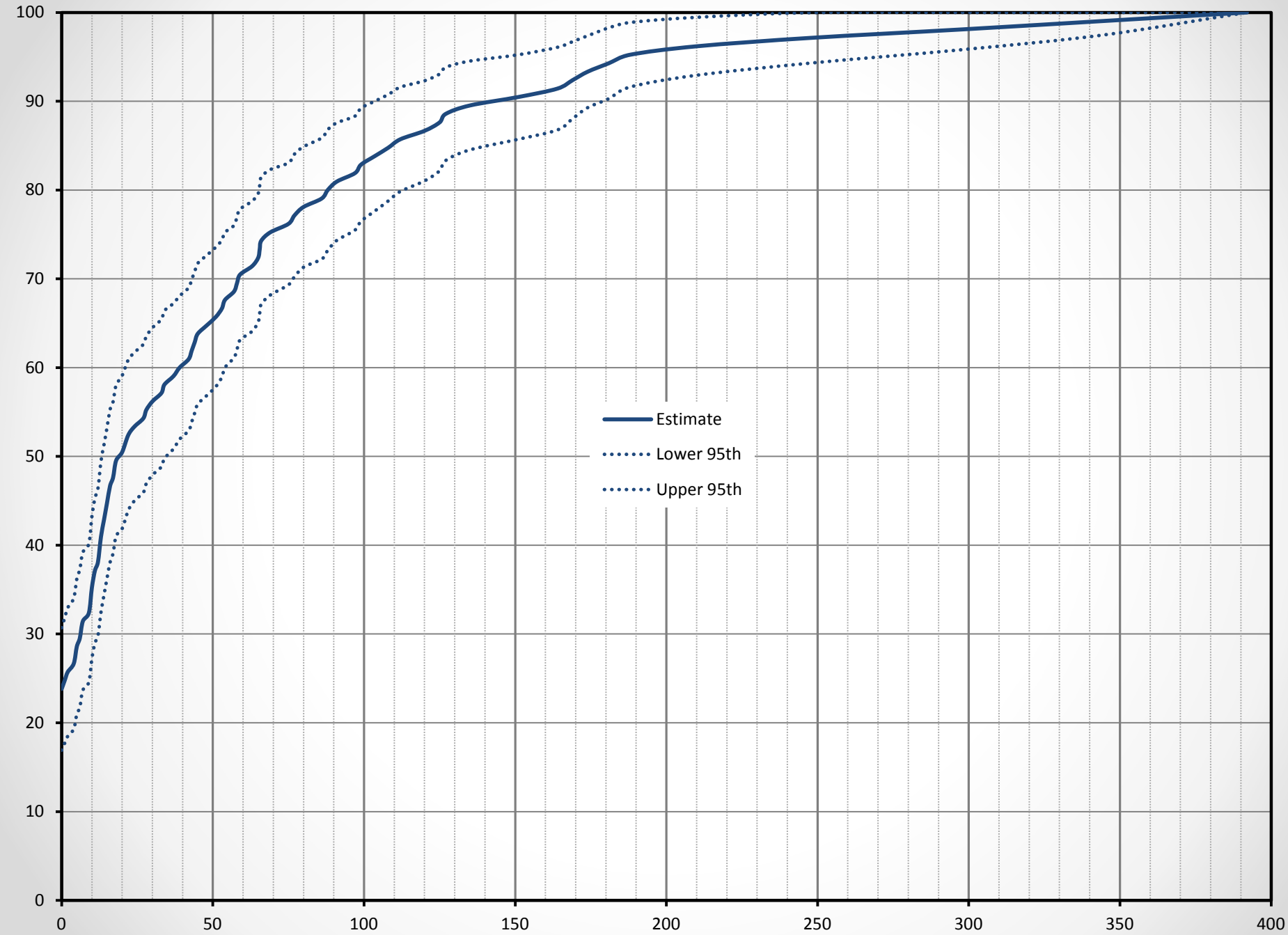


Log Normal Distribution



What/Who is normal?




100
90
80
70
60
50
40
30
20
10
0
0
50
100
150
200
250
300
350
400
Estimate
Lower 95th
Upper 95th

The Statistics

Distribution #1

	ALL	FCR >0 Only
Median	17.5	39.0
Mean	47.5	63.3
90	127.8	158.8
95	174.7	194.6
99	298.5	310.2

Distribution #2

	ALL	FCR >0 Only
Median	17.5	23.0
Mean	47.5	52.8
90	127.8	136.6
95	174.7	181.6
99	298.5	303.2

The Statistics

Distribution #1

	ALL	FCR >0 Only
Median	17.5	**39.0**
Mean	47.5	**63.3**
90	127.8	**158.8**
95	174.7	**194.6**
99	298.5	**310.2**

Distribution #2

	ALL	FCR >0 Only
Median	17.5	**23.0**
Mean	47.5	**52.8**
90	127.8	**136.6**
95	174.7	**181.6**
99	298.5	**303.2**

The Statistics

Distribution #1

	ALL	FCR >0 Only
Median	17.5	39.0
Mean	47.5	63.3
90	127.8	158.8
95	174.7	194.6
99	298.5	310.2

Distribution #2

	ALL	FCR >0 Only
Median	17.5	23.0
Mean	47.5	52.8
90	127.8	136.6
95	174.7	181.6
99	298.5	303.2

The Statistics

Distribution #1

	ALL	>0 Only
Median	17.5	39.0
Mean	47.5	63.3
90	127.8	158.8
95	174.7	194.6
99	298.5	310.2

Distribution #3

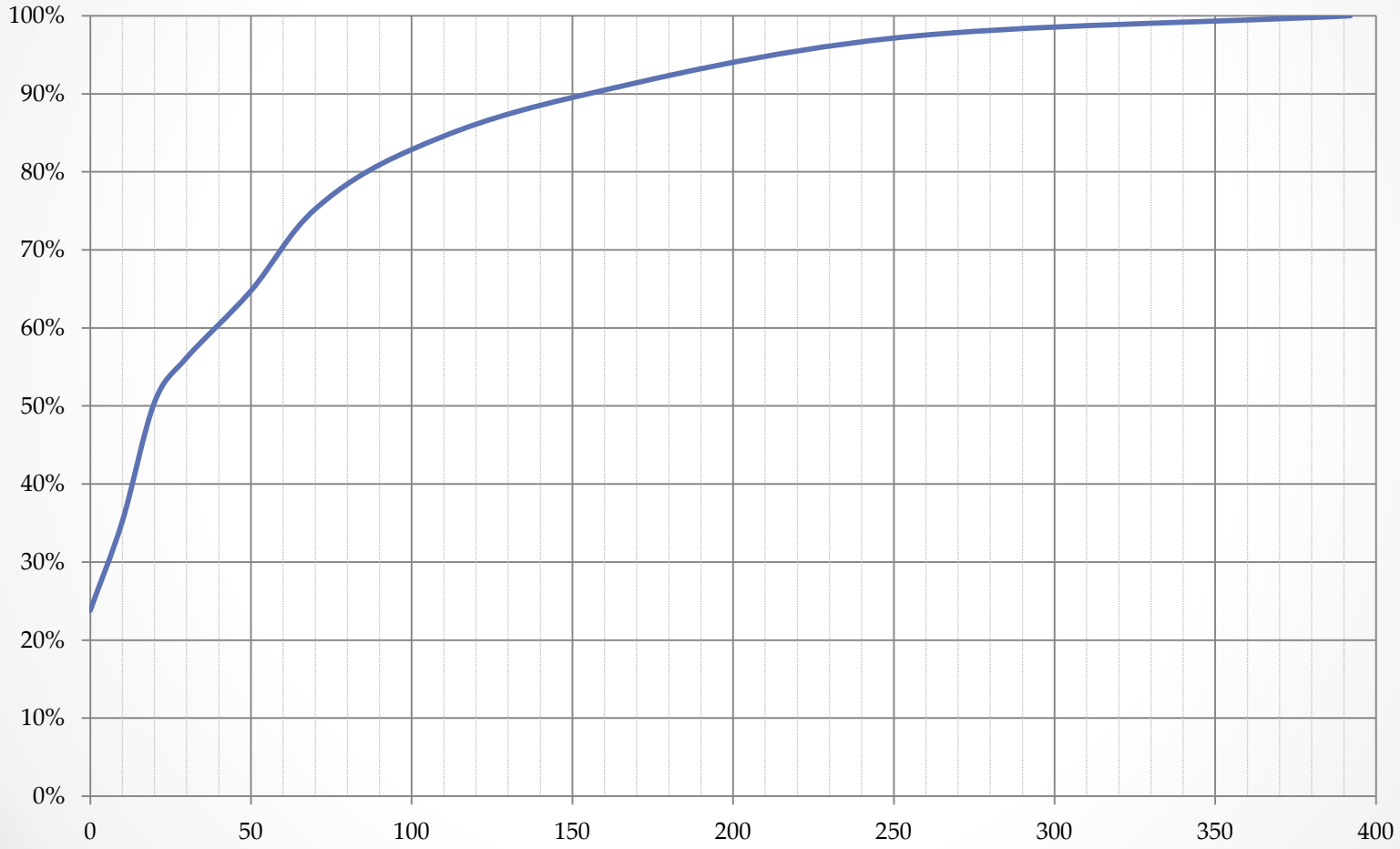
	ALL	>0 Only	High
Median	20.0	43.5	103.5
Mean	52.7	69.2	157.2
90	144.6	169.5	307.6
95	186.6	211.9	349.8
99	343.1	354.9	383.6

Summary

- Misidentifying low consumers as non-consumers will raise consumption rate statistical estimates, create a high bias
- This matters only IF we don't use statistics based on the whole population
- Missing a fraction of the population who are high end consumers will lower consumption rate statistical estimates, create a low bias
- If we have an idea of how many we missed and how much they eat, we can construct a fuller distribution
- Biases due to these sampling issues are likely small compared to the range in choice of statistic

Questions?

Cumulative Density Plot



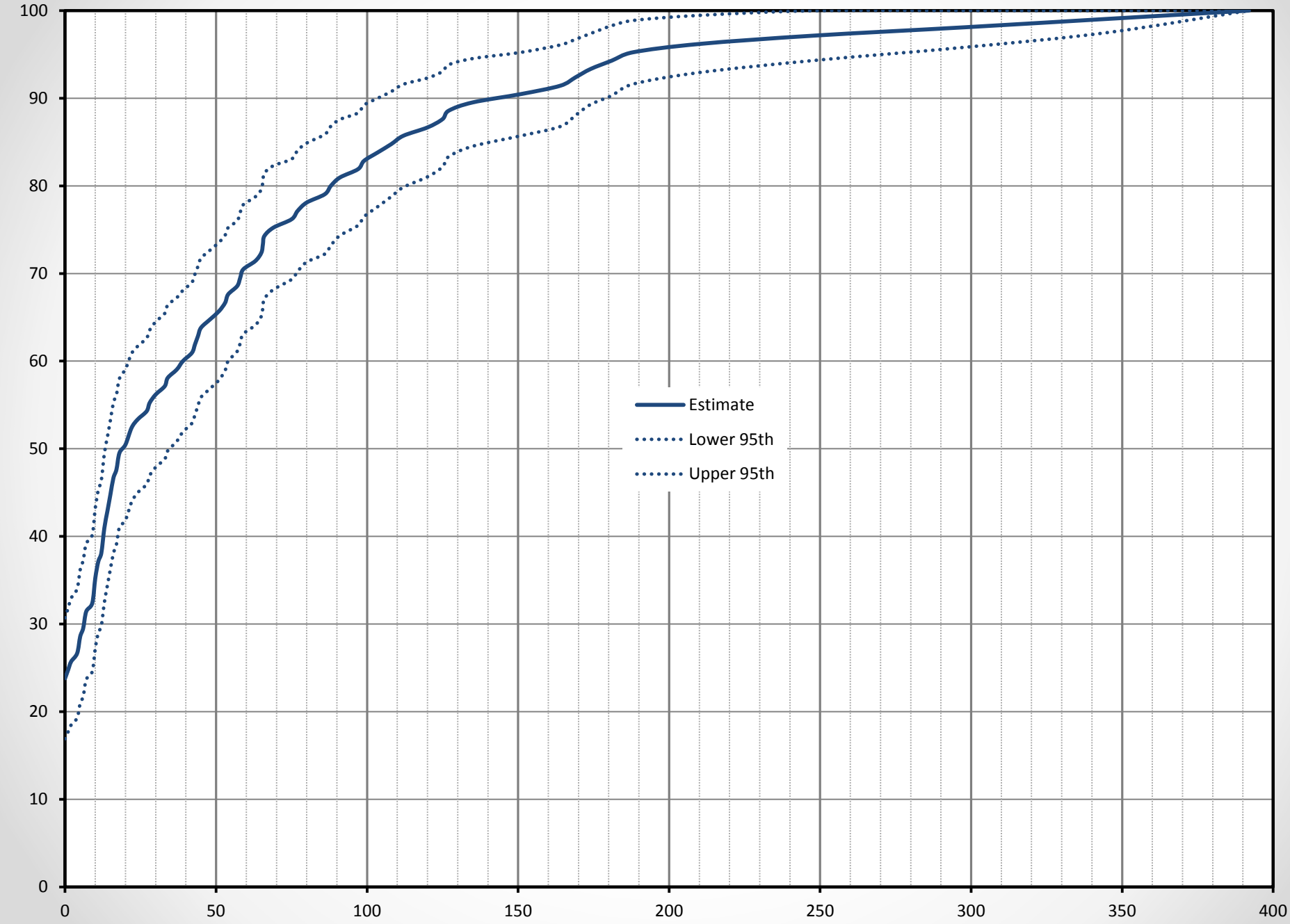
The Statistics

Distribution #1

	ALL	>0 Only
Mean	47.5	63.3
Median	17.5	39.0
90	127.8	158.8
95	174.7	194.6
99	298.5	310.2

Distribution #3

	ALL	>0 Only	High
Mean	52.7	69.2	157.2
Median	20.0	43.5	103.5
90	144.6	169.5	307.6
95	186.6	211.9	349.8
99	343.1	354.9	383.6


100
90
80
70
60
50
40
30
20
10
0
0
50
100
150
200
250
300
350
400
Estimate
Lower 95th
Upper 95th