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
Freshwater/Estuarine Finfish and Shellfish

Number of Persons (millions)

Finfish and Shellfish Consumed (grams/person/day)
Hypothetical fish consumption distribution #1

Consumption Rate, g/day

Frequency

0 10 20 30 50 70 150 250 More
Hypothetical fish consumption distribution #2
Hypothetical fish consumption distribution #3

Frequency

0 10 20 30 50 70 100 150 250 More

Consumption Rate, g/day
Hypothetical fish consumption distribution #1

Frequency

Consumption Rate, g/day

0 10 20 30 50 70 100 150 250 More
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 50\textsuperscript{th} percentile
  2. Mean
  3. 90\textsuperscript{th} percentile
  4. 95\textsuperscript{th} percentile
  5. 99\textsuperscript{th} percentile
Log Normal Distribution

![Histogram of Egg to Smolt Survival](image)
What/Who is normal?
# The Statistics

## Distribution #1

<table>
<thead>
<tr>
<th></th>
<th>ALL</th>
<th>FCR &gt;0 Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>17.5</td>
<td>39.0</td>
</tr>
<tr>
<td>Mean</td>
<td>47.5</td>
<td>63.3</td>
</tr>
<tr>
<td>90</td>
<td>127.8</td>
<td>158.8</td>
</tr>
<tr>
<td>95</td>
<td>174.7</td>
<td>194.6</td>
</tr>
<tr>
<td>99</td>
<td>298.5</td>
<td>310.2</td>
</tr>
</tbody>
</table>

## Distribution #2

<table>
<thead>
<tr>
<th></th>
<th>ALL</th>
<th>FCR &gt;0 Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>17.5</td>
<td>23.0</td>
</tr>
<tr>
<td>Mean</td>
<td>47.5</td>
<td>52.8</td>
</tr>
<tr>
<td>90</td>
<td>127.8</td>
<td>136.6</td>
</tr>
<tr>
<td>95</td>
<td>174.7</td>
<td>181.6</td>
</tr>
<tr>
<td>99</td>
<td>298.5</td>
<td>303.2</td>
</tr>
</tbody>
</table>
# The Statistics

## Distribution #1

<table>
<thead>
<tr>
<th></th>
<th>ALL</th>
<th>FCR &gt;0 Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>17.5</td>
<td>39.0</td>
</tr>
<tr>
<td>Mean</td>
<td>47.5</td>
<td>63.3</td>
</tr>
<tr>
<td>90</td>
<td>127.8</td>
<td>158.8</td>
</tr>
<tr>
<td>95</td>
<td>174.7</td>
<td>194.6</td>
</tr>
<tr>
<td>99</td>
<td>298.5</td>
<td>310.2</td>
</tr>
</tbody>
</table>

## Distribution #2

<table>
<thead>
<tr>
<th></th>
<th>ALL</th>
<th>FCR &gt;0 Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>17.5</td>
<td>23.0</td>
</tr>
<tr>
<td>Mean</td>
<td>47.5</td>
<td>52.8</td>
</tr>
<tr>
<td>90</td>
<td>127.8</td>
<td>136.6</td>
</tr>
<tr>
<td>95</td>
<td>174.7</td>
<td>181.6</td>
</tr>
<tr>
<td>99</td>
<td>298.5</td>
<td>303.2</td>
</tr>
</tbody>
</table>
# The Statistics

## Distribution #1

<table>
<thead>
<tr>
<th></th>
<th>ALL</th>
<th>FCR &gt;0 Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>17.5</td>
<td>39.0</td>
</tr>
<tr>
<td>Mean</td>
<td>47.5</td>
<td>63.3</td>
</tr>
<tr>
<td>90</td>
<td>127.8</td>
<td>158.8</td>
</tr>
<tr>
<td>95</td>
<td>174.7</td>
<td>194.6</td>
</tr>
<tr>
<td>99</td>
<td>298.5</td>
<td>310.2</td>
</tr>
</tbody>
</table>

## Distribution #2

<table>
<thead>
<tr>
<th></th>
<th>ALL</th>
<th>FCR &gt;0 Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>17.5</td>
<td>23.0</td>
</tr>
<tr>
<td>Mean</td>
<td>47.5</td>
<td>52.8</td>
</tr>
<tr>
<td>90</td>
<td>127.8</td>
<td>136.6</td>
</tr>
<tr>
<td>95</td>
<td>174.7</td>
<td>181.6</td>
</tr>
<tr>
<td>99</td>
<td>298.5</td>
<td>303.2</td>
</tr>
</tbody>
</table>
The Statistics

<table>
<thead>
<tr>
<th></th>
<th>Distribution #1</th>
<th>Distribution #3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ALL</td>
<td>&gt;0 Only</td>
</tr>
<tr>
<td>Median</td>
<td>17.5</td>
<td>39.0</td>
</tr>
<tr>
<td>Mean</td>
<td>47.5</td>
<td>63.3</td>
</tr>
<tr>
<td>90</td>
<td>127.8</td>
<td>158.8</td>
</tr>
<tr>
<td>95</td>
<td>174.7</td>
<td>194.6</td>
</tr>
<tr>
<td>99</td>
<td>298.5</td>
<td>310.2</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th></th>
<th>ALL</th>
<th>&gt;0 Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>47.5</td>
<td>63.3</td>
</tr>
<tr>
<td>Median</td>
<td>17.5</td>
<td>39.0</td>
</tr>
<tr>
<td>90</td>
<td>127.8</td>
<td>158.8</td>
</tr>
<tr>
<td>95</td>
<td>174.7</td>
<td>194.6</td>
</tr>
<tr>
<td>99</td>
<td>298.5</td>
<td>310.2</td>
</tr>
</tbody>
</table>

### Distribution #3

<table>
<thead>
<tr>
<th></th>
<th>ALL</th>
<th>&gt;0 Only</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>52.7</td>
<td>69.2</td>
<td>157.2</td>
</tr>
<tr>
<td>Median</td>
<td>20.0</td>
<td>43.5</td>
<td>103.5</td>
</tr>
<tr>
<td>90</td>
<td>144.6</td>
<td>169.5</td>
<td>307.6</td>
</tr>
<tr>
<td>95</td>
<td>186.6</td>
<td>211.9</td>
<td>349.8</td>
</tr>
<tr>
<td>99</td>
<td>343.1</td>
<td>354.9</td>
<td>383.6</td>
</tr>
</tbody>
</table>