DISTRIBUTIONS

As we saw last time, a well-drawn graph conveys a lot of useful information...

but a poorly drawn graph can mislead and confuse.

We would like a quantitative method of describing distributions may not entirely avoid misinformation, but at least the limitations will be identifiable.

TABLE FORMAT

<table>
<thead>
<tr>
<th>Exact Limits</th>
<th>Midpoint</th>
<th>f</th>
<th>cf</th>
<th>%</th>
<th>c%</th>
</tr>
</thead>
<tbody>
<tr>
<td>64.5–69.5</td>
<td>67</td>
<td>6</td>
<td>180</td>
<td>3.33</td>
<td>100</td>
</tr>
<tr>
<td>59.5–64.5</td>
<td>62</td>
<td>15</td>
<td>174</td>
<td>8.33</td>
<td>96.67</td>
</tr>
<tr>
<td>54.5–59.5</td>
<td>57</td>
<td>37</td>
<td>159</td>
<td>20.56</td>
<td>88.34</td>
</tr>
<tr>
<td>49.5–54.5</td>
<td>52</td>
<td>30</td>
<td>122</td>
<td>16.67</td>
<td>77.78</td>
</tr>
<tr>
<td>44.5–49.5</td>
<td>47</td>
<td>42</td>
<td>92</td>
<td>23.33</td>
<td>51.11</td>
</tr>
<tr>
<td>39.5–44.5</td>
<td>42</td>
<td>22</td>
<td>50</td>
<td>12.22</td>
<td>27.78</td>
</tr>
<tr>
<td>34.5–39.5</td>
<td>37</td>
<td>18</td>
<td>28</td>
<td>10.00</td>
<td>15.56</td>
</tr>
<tr>
<td>29.5–34.5</td>
<td>32</td>
<td>7</td>
<td>10</td>
<td>3.89</td>
<td>5.56</td>
</tr>
<tr>
<td>24.5–29.5</td>
<td>27</td>
<td>2</td>
<td>3</td>
<td>1.11</td>
<td>1.07</td>
</tr>
<tr>
<td>19.5–24.5</td>
<td>22</td>
<td>1</td>
<td>1</td>
<td>0.56</td>
<td>0.56</td>
</tr>
</tbody>
</table>

FREQUENCY DISTRIBUTIONS

frequency versus score class interval

DISTRIBUTION USES

summarize data
indicate most frequent data values
indicate amount of variation across data values
allows us to interpret a single score in the context of other scores
we will explore quantitative methods to describe distributions
PERCENTILE

Point in a distribution at (or below) which a given percentage of scores is found.

Written as $P_{\text{percentage}}$.

28th percentile is written as $P_{28}$.

99th percentile is written as $P_{99}$.

...
CALCULATIONS

\[ P_X = ll + \left( \frac{np - cf}{f_i} \right) (w) \]

\( ll \) = exact lower limit of the interval containing the percentile point
\( n \) = total number of scores
\( p \) = \( X/100 \), proportion corresponding to percentile (decimal form)
\( cf \) = cumulative frequency of scores below the interval containing the percentile point
\( f_i \) = frequency of scores in the interval containing the percentile point
\( w \) = width of class interval

PERCENTILE RANK

given a particular data value, what percentage of data values are smaller?
e.g. given a score on a test, what percentage of scores were lower?
sort of the reverse of percentile
for a data value of 39, we write the percentile rank as \( PR_{39} \)
(Used on achievement tests!)

\[ PR_X = \left( \frac{cf + (f_i)(X - ll)/w}{n} \right) (100) \]

\[ PR_{39} = \frac{10 + (18)(39 - 34.5)/5}{180} (100) \]

Exact Limits | Midpoint | \( f \) | cf | % | c%
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>34.5-69.5</td>
<td>67</td>
<td>6</td>
<td>180</td>
<td>3.33</td>
<td>100</td>
</tr>
<tr>
<td>39.5-64.5</td>
<td>62</td>
<td>15</td>
<td>174</td>
<td>8.33</td>
<td>90.67</td>
</tr>
<tr>
<td>34.5-59.5</td>
<td>57</td>
<td>37</td>
<td>159</td>
<td>20.56</td>
<td>88.34</td>
</tr>
<tr>
<td>49.5-54.5</td>
<td>52</td>
<td>30</td>
<td>122</td>
<td>16.67</td>
<td>67.78</td>
</tr>
<tr>
<td>44.5-49.5</td>
<td>47</td>
<td>42</td>
<td>92</td>
<td>23.33</td>
<td>51.11</td>
</tr>
<tr>
<td>39.5-44.5</td>
<td>42</td>
<td>22</td>
<td>50</td>
<td>12.22</td>
<td>27.78</td>
</tr>
<tr>
<td>34.5-39.5</td>
<td>37</td>
<td>18</td>
<td>28</td>
<td>10.00</td>
<td>15.56</td>
</tr>
<tr>
<td>29.5-34.5</td>
<td>32</td>
<td>10</td>
<td>10</td>
<td>3.89</td>
<td>5.56</td>
</tr>
<tr>
<td>24.5-29.5</td>
<td>27</td>
<td>2</td>
<td>3</td>
<td>1.11</td>
<td>1.67</td>
</tr>
<tr>
<td>19.5-24.5</td>
<td>22</td>
<td>1</td>
<td>1</td>
<td>0.56</td>
<td>0.56</td>
</tr>
</tbody>
</table>

LIMITATIONS

percentiles help describe a data value relative to its frequency distribution
but they have some drawbacks
percentiles use an ordinal scale
equal differences in percentiles do not indicate equal differences in raw scores!
class intervals with higher frequency cover a broader range of percentiles
(steeper part of ogive)
LIMITATIONS

percentiles exaggerate differences in scores when lots of people have similar scores
underestimate actual differences when lots of people have very different scores
differences in percentiles should not be compared across different distributions!!!
(only provide information on relative ranking of scores: ordinal scale!)
cannot be meaningfully averaged, summed, multiplied....
fixing these problems requires additional terms for describing distributions (central tendency)

CONCLUSIONS

percentiles
percentile ranks

NEXT TIME

central tendency (mode, median, mean)
variation
Do you deserve a pay raise?