Introduction to Statistics in Psychology
PSY 201
Professor Greg Francis
Lecture 05
central tendency

Does a company deserve a tax break?

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 are exploring quantitative methods to describe distributions

LIMITATIONS
Last time we discussed percentiles and percentile ranks
very useful for comparing a score to a distribution of scores
not so good for talking about a distribution overall
want to quantify ideas of central tendency (most of scores, average score,...)
• mode
• median
• mean

and variation (how variable scores are in a distribution)

MODE
the most frequent data value (score)
easy to find from a table of frequency scores

<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>67.88</td>
</tr>
<tr>
<td>44.5–49.5</td>
<td>47</td>
<td>22</td>
<td>42</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>1.67</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>

we actually look for a modal interval and consider the midpoint of the interval to be the mode

unimodal distribution: when there is a single mode. (single hill)
multimodal distribution: when there are several modes. (many hills)
bimodal distribution: when there are two modes. (two hills)

NOTE: the use of the terms are not quite consistent!
BIMODAL

this distribution might be called bimodal, even though there is really only one mode!

not very useful for mathematics!

MEDIAN

the point below which 50% of scores fall

the 50th percentile

\[ \text{Mdn} = P_{50} = I + \left( \frac{n(0.5) - cf}{f_i} \right)(w) \]

CALCULATIONS

when the raw scores are used (instead of class intervals)

1. Arrange the scores in ascending order (from lowest to highest).
2. If there is an odd number of scores, the median is the middle score.
3. If there is an even number of scores the median is halfway between the two middle scores.

Will this always give the same value as for the frequency distribution approach?

scores: 89, 92, 94, 95 (even number of scores)
the median is: halfway between 92 and 94 = 93

scores: 83, 89, 92, 94, 95 (odd number of scores)
the median is: the middle score = 92

MEAN

arithmetic average of scores in a distribution

mean of a population is designated as \( \mu \)
mean of a sample is designated as \( \bar{X} \)

Calculated as:

\[ \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \]

\[ X_i = \text{the } i\text{th score} \]
\[ n = \text{total number of scores} \]

sometimes just written as

\[ \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \]
CALCULATIONS

<table>
<thead>
<tr>
<th>Name</th>
<th>Sex</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greg</td>
<td>Male</td>
<td>95</td>
</tr>
<tr>
<td>Ian</td>
<td>Male</td>
<td>89</td>
</tr>
<tr>
<td>Aimee</td>
<td>Female</td>
<td>94</td>
</tr>
<tr>
<td>Jim</td>
<td>Male</td>
<td>92</td>
</tr>
</tbody>
</table>

\[ X = \frac{1}{n} \sum X_i \]

\[ X = \frac{1}{4} [X_1 + X_2 + X_3 + X_4] \]

\[ X = \frac{1}{4} [95 + 89 + 94 + 92] = \frac{370}{4} = 92.5 \]

COMPARISON

- mean can only be used on interval or ratio data.
- mode can be used on nominal data
- mode and median can be used on ordinal data
- mean can be manipulated mathematically
- mean can be sensitive to extreme scores

TAX BREAKS

<table>
<thead>
<tr>
<th>Position</th>
<th>Number</th>
<th>Salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>President</td>
<td>1</td>
<td>154,000</td>
</tr>
<tr>
<td>Ex. vice pres.</td>
<td>1</td>
<td>140,000</td>
</tr>
<tr>
<td>Vice pres.</td>
<td>2</td>
<td>120,000</td>
</tr>
<tr>
<td>Controller</td>
<td>1</td>
<td>92,800</td>
</tr>
<tr>
<td>Senior sales</td>
<td>3</td>
<td>80,000</td>
</tr>
<tr>
<td>Junior sales</td>
<td>4</td>
<td>42,800</td>
</tr>
<tr>
<td>Foreman</td>
<td>1</td>
<td>37,000</td>
</tr>
<tr>
<td>Machinists</td>
<td>12</td>
<td>25,000</td>
</tr>
</tbody>
</table>

The company wants a tax break from the city. Is it a good corporate citizen?

Mean = $67,640
Median = $37,000
Mode = $25,000

MEAN OF MEANS

<table>
<thead>
<tr>
<th>Name</th>
<th>Sex</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greg</td>
<td>Male</td>
<td>95</td>
</tr>
<tr>
<td>Ian</td>
<td>Male</td>
<td>89</td>
</tr>
<tr>
<td>Aimee</td>
<td>Female</td>
<td>94</td>
</tr>
<tr>
<td>Jim</td>
<td>Male</td>
<td>92</td>
</tr>
</tbody>
</table>

\[ \bar{X} = \frac{\sum X_i}{n} \]

\[ \bar{X} = \frac{X_1 + X_2 + X_3 + X_4}{4} \]

\[ \bar{X} = \frac{95 + 89 + 94 + 92}{4} = \frac{370}{4} = 92.5 \]

COMBINED GROUPS

the mean of means is not the same thing as the mean of all the scores in all the groups!!!

Consider our small data set

\[ X_{F} = \frac{94}{1} = 94.0 \]

\[ X_{M} = \frac{95 + 89 + 92}{3} = 92.0 \]

the mean of the means is:

\[ \frac{X_{F} + X_{M}}{2} = \frac{94 + 92}{2} = 93.0 \]

but we already found that the mean of all the scores was

\[ \bar{X} = 92.5 \]

too much weight on the “female” group

correct calculation goes like

\[ \bar{X} = \frac{n_{F}X_{F} + n_{M}X_{M}}{n_{F} + n_{M}} \]

where

\[ X_{F} \] is the mean for the females

\[ X_{M} \] is the mean for the males

\[ n_{F} \] is the number of females

\[ n_{M} \] is the number of males

\[ \bar{X} = \frac{(1)(94.0) + (3)(92.0)}{1 + 3} = \frac{94 + 276}{4} = 92.5 \]

same as direct calculation of \( \bar{X} \)!
COMBINED GROUPS

in general, given

\[ \overline{X}_i = \text{individual group means} \]
\[ n_i = \text{number of observations in individual groups} \]
\[ N = \sum n_i = \text{total number of observations in all groups} \]

\[ \overline{X} = \frac{\sum n_i \overline{X}_i}{N} \]

PROPERTIES OF MEAN

1. The sum of deviations of all scores from the mean is zero.

2. The sum of squares of the deviations from the mean is smaller than the sum of squares of deviations from any other value.

\[ x_i = X_i - \overline{X} \]

pluses and minuses cancel each other out!

\[ \sum x_i = \sum (X_i - \overline{X}) \]
\[ = (95 - 92.5) + (89 - 92.5) + (94 - 92.5) + (92 - 92.5) \]
\[ = 2.5 + (-3.5) + (1.5) + (-0.5) = 0 \]

SUM OF SQUARES

we ignore the direction of deviation, and consider the squared magnitude of deviation

\[ \sum x_i^2 = \sum (X_i - \overline{X})^2 \]
\[ = (95 - 92.5)^2 + (89 - 92.5)^2 + (94 - 92.5)^2 + (92 - 92.5)^2 \]
\[ = 2.5^2 + (-3.5)^2 + (1.5)^2 + (-0.5)^2 \]
\[ = 6.25 + 12.25 + 2.25 + 0.25 = 21.0 \]

sum of squared deviations from any other value is larger

\[ \sum (X_i - 90)^2 = 25 + 1 + 16 + 4 = 46.0 \]
\[ \sum (X_i - 100)^2 = 25 + 121 + 36 + 64 = 246.0 \]

these properties will be important later!

CONCLUSIONS

central tendency

mode
mean
median

NEXT TIME

variation
variance
standard deviation
z scores

How to make IQ scores look good.