Describing Location in a Distribution

2.1 Measures of Relative Standing and Density Curves

YMS3e

AP Stats at LSHS
Mr. Molesky
Sample Data

Consider the following test scores for a small class:

<table>
<thead>
<tr>
<th>79</th>
<th>81</th>
<th>80</th>
<th>77</th>
<th>73</th>
<th>83</th>
<th>74</th>
<th>93</th>
<th>78</th>
<th>80</th>
<th>75</th>
<th>67</th>
<th>73</th>
</tr>
</thead>
<tbody>
<tr>
<td>77</td>
<td>83</td>
<td>86</td>
<td>90</td>
<td>79</td>
<td>85</td>
<td>83</td>
<td>89</td>
<td>84</td>
<td>82</td>
<td>77</td>
<td>72</td>
<td></td>
</tr>
</tbody>
</table>

Jenny’s score is noted in red. How did she perform on this test relative to her peers?

Her score is “above average”... but how far above average is it?
One way to describe relative position in a data set is to tell how many standard deviations above or below the mean the observation is.

**Standardized Value: “z-score”**

If the mean and standard deviation of a distribution are known, the “z-score” of a particular observation, \( x \), is:

\[
z = \frac{x - \text{mean}}{\text{standard deviation}}
\]
Calculating z-scores

- Consider the test data and Julia’s score.

<table>
<thead>
<tr>
<th>79</th>
<th>81</th>
<th>80</th>
<th>77</th>
<th>73</th>
<th>83</th>
<th>74</th>
<th>93</th>
<th>78</th>
<th>80</th>
<th>75</th>
<th>67</th>
<th>73</th>
</tr>
</thead>
<tbody>
<tr>
<td>77</td>
<td>83</td>
<td>86</td>
<td>90</td>
<td>79</td>
<td>85</td>
<td>83</td>
<td>89</td>
<td>84</td>
<td>82</td>
<td>77</td>
<td>72</td>
<td></td>
</tr>
</tbody>
</table>

According to Minitab, the mean test score was 80 while the standard deviation was 6.07 points.

Julia’s score was above average. Her standardized z-score is:

\[ z = \frac{x - 80}{6.07} = \frac{86 - 80}{6.07} = 0.99 \]

Julia’s score was almost one full standard deviation above the mean. What about Kevin: x=
## Calculating z-scores

<table>
<thead>
<tr>
<th>79</th>
<th>81</th>
<th>80</th>
<th>77</th>
<th>73</th>
<th>83</th>
<th>74</th>
<th>93</th>
<th>78</th>
<th>80</th>
<th>75</th>
<th>67</th>
<th>73</th>
</tr>
</thead>
<tbody>
<tr>
<td>77</td>
<td>83</td>
<td><strong>86</strong></td>
<td>90</td>
<td>79</td>
<td>85</td>
<td>83</td>
<td>89</td>
<td>84</td>
<td>82</td>
<td>77</td>
<td>72</td>
<td></td>
</tr>
</tbody>
</table>

### Julia

Julia: $z = \frac{(86 - 80)}{6.07}$

$z = 0.99$ \{above average = $+z$\}

### Kevin

Kevin: $z = \frac{(72 - 80)}{6.07}$

$z = -1.32$ \{below average = $-z$\}

### Katie

Katie: $z = \frac{(80 - 80)}{6.07}$

$z = 0$ \{average $z = 0$\}
Comparing Scores

- Standardized values can be used to compare scores from two different distributions.
- Statistics Test: mean = 80, std dev = 6.07
- Chemistry Test: mean = 76, std dev = 4
- Jenny got an 86 in Statistics and 82 in Chemistry.
- On which test did she perform better?

\[
\text{Statistics: } z = \frac{86 - 80}{6.07} = 0.99 \\
\text{Chemistry: } z = \frac{82 - 76}{4} = 1.5
\]

Although she had a lower score, she performed relatively better in Chemistry.
Percentiles

Another measure of relative standing is a percentile rank.

\textbf{p}^{th} \textbf{percentile}: Value with p \% of observations below it.

- median = 50th percentile \quad \text{(mean=50th \%ile if symmetric)}
- Q1 = 25th percentile
- Q3 = 75th percentile

Jenny got an 86.
22 of the 25 scores are \leq 86.
Jenny is in the $\frac{22}{25} = 88$th \%ile.
Chebyshev’s Inequality

- The % of observations at or below a particular z-score depends on the shape of the distribution.

- An interesting (non-AP topic) observation regarding the % of observations around the mean in ANY distribution is Chebyshev’s Inequality.

**Chebyshev’s Inequality:**

In any distribution, the % of observations within $k$ standard deviations of the mean is at least

\[
\% \text{ within } k \text{ std dev } \geq 1 - \frac{1}{k^2}
\]
Density Curve

- In Chapter 1, you learned how to plot a dataset to describe its shape, center, spread, etc.
- *Sometimes, the overall pattern of a large number of observations is so regular that we can describe it using a smooth curve.*

**Density Curve:**
An idealized description of the overall pattern of a distribution.
Area underneath = 1, representing 100% of observations.
Density Curves come in many different shapes; symmetric, skewed, uniform, etc.
The area of a region of a density curve represents the % of observations that fall in that region.
The median of a density curve cuts the area in half.
The mean of a density curve is its “balance point.”
2.1 Summary

- We can describe the overall pattern of a distribution using a density curve.
- The area under any density curve = 1. This represents 100% of observations.
- Areas on a density curve represent % of observations over certain regions.
- An individual observation’s relative standing can be described using a z-score or percentile rank.

\[ z = \frac{x - \text{mean}}{\text{standard deviation}} \]