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:

79	81	80	77	73	83	74	93	78	80	75	67	73
77	83	86	90	79	85	83	89	84	82	77	72	

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

```
6 | 7
7 | 2334
7 | 5777899
8 | 00123334
8 | 569
9 | 03
```

Minitab

Session

Descriptive Statistics: Test 1 scores

```
Variable        N    Mean   Median  TrMean  StDev  SE Mean
Test 1 scores   25   80.00  80.00   80.00   6.07   1.21

Variable        Minimum  Maximum  Q1     Q3
Test 1 scores   67.00    93.00    76.00  83.50
```

Her score is "above average"… but how far above average is it?

Standardized Value

- **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.**

<u>Standardized Value: "z-score"</u>

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.

79	**81**	**80**	**77**	**73**	**83**	**74**	**93**	**78**	**80**	**75**	**67**	**73**
77	**83**	**86**	**90**	**79**	**85**	**83**	**89**	**84**	**82**	**77**	**72**	

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

79	81	80	77	73	83	74	93	78	80	75	67	73
77	83	86	90	79	85	83	89	84	82	77	72	

6	7
7	2334
7	5777899
8	00123334
8	569
9	03

Julia: z=(86-80) / 6.07

z= 0.99 *{above average = +z}*

Kevin: z=(72-80) / 6.07

z= -1.32 *{below average = -z}*

Katie: z=(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?

Statistics

$$z = \frac{86 - 80}{6.07} = 0.99$$

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**.
- **p^{th} percentile**: Value with p % of observations below it.
  - median = 50th percentile {*mean=50th %ile if symmetric*}
  - Q1 = 25th percentile
  - Q3 = 75th percentile

Jenny got an 86.
22 of the 25 scores are ≤ 86.
Jenny is in the 22/25 = 88th %ile.

6	7
7	2334
7	5777899
8	00123334
8	569
9	03

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*

$$\%\text{within } k \text{ std dev} \geq \left(1 - \frac{1}{k^2}\right)$$

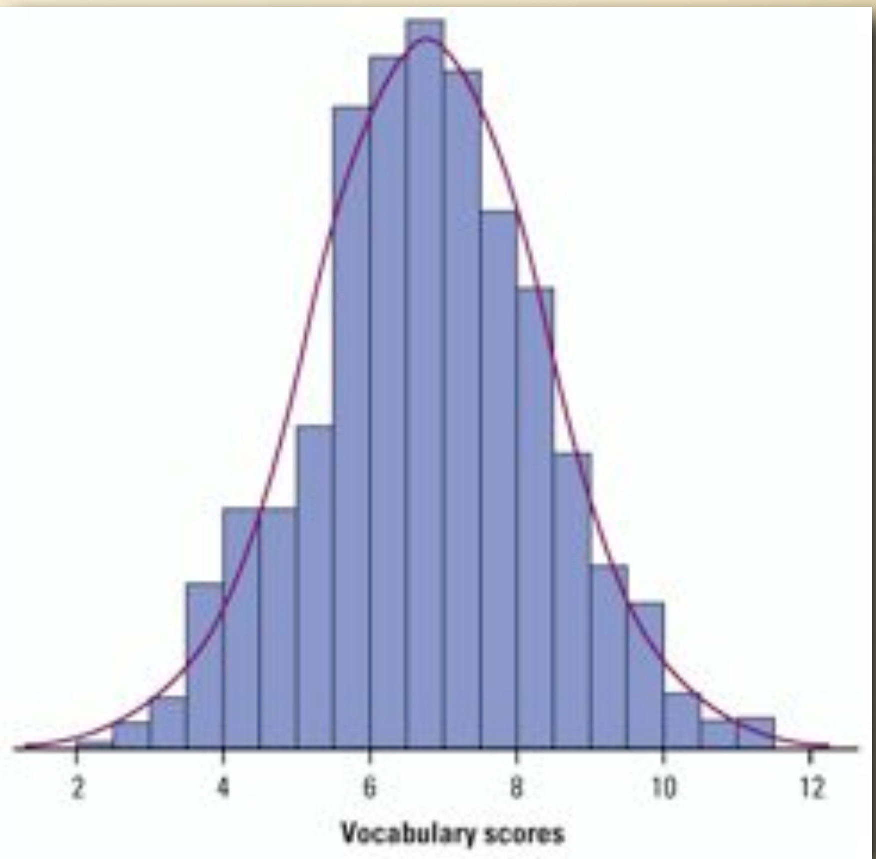
Density Curve

- 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:

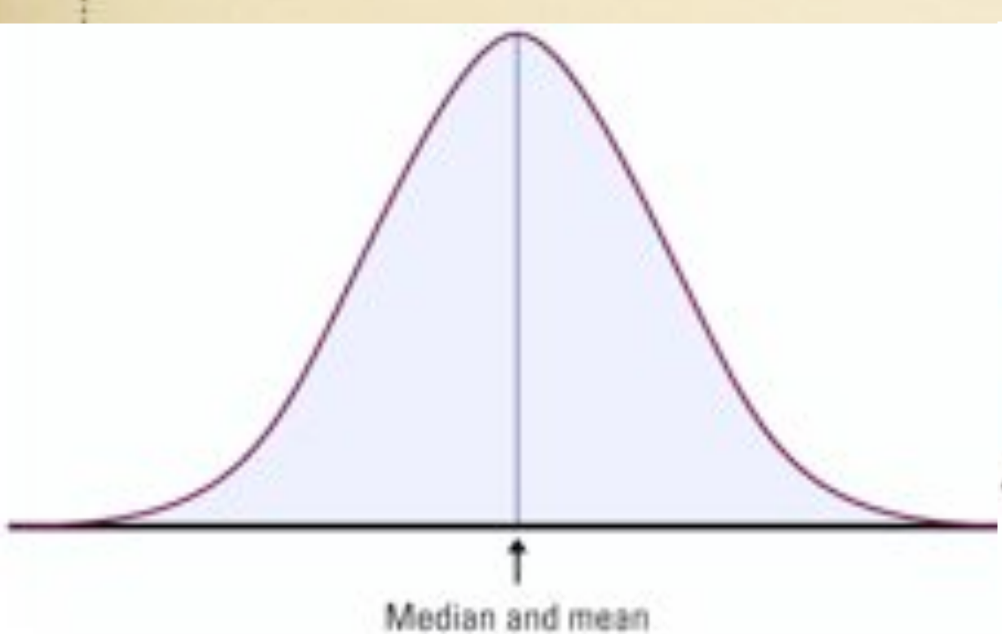
An idealized description of the overall pattern of a distribution.

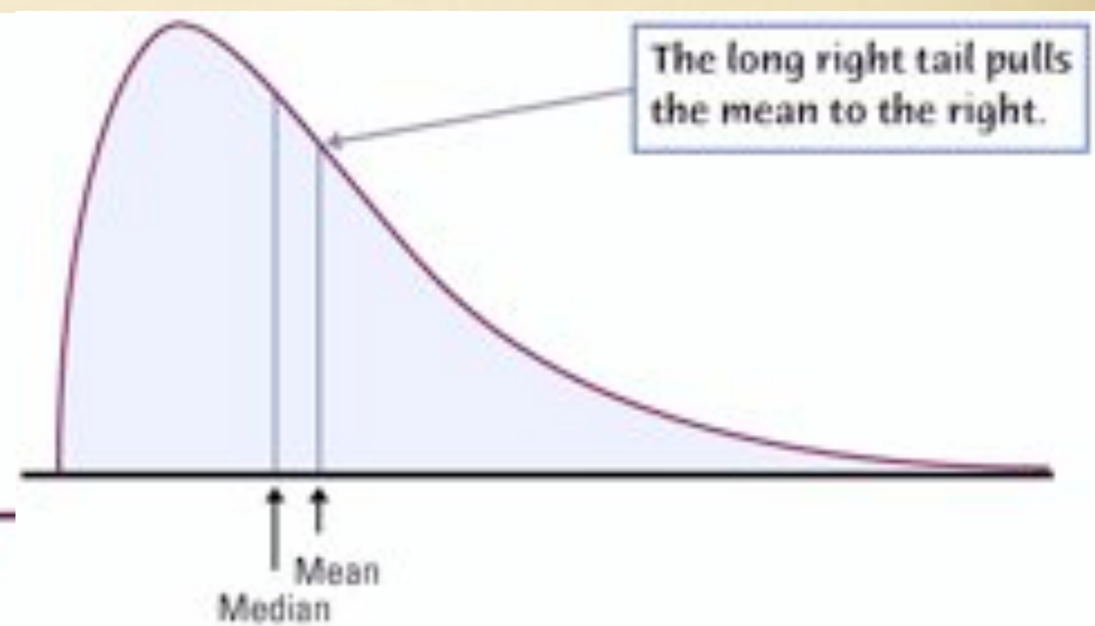
Area underneath = 1, representing 100% of observations.



Density Curves

- 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**.

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