**Activity 7.2.3 – Examining Values in Data Distributions**

**Common Shapes of Data Distributions**



 Bell-Shaped Skewed-Right Skewed-Left Uniform

Data distributions follow different shapes depending on how their values are distributed.

* Bell-shaped distributions have data that are approximately symmetrically distributed around the mean where the mean, median and mode are all equal. These distributions are called *approximately normal*.
* Skewed distributions are unbalanced distributions that have a small number of very low or very high data values.
* Uniform distributions have data that are evenly distributed among all data values.

**Empirical Rule for Approximately Normal Distributions**

The Empirical Rule describes the variability of data values in a data set based on the data set’s mean and standard deviation.

$$μ-3σ μ-2σ μ-σ μ μ+σ μ+2σ μ+3σ$$

Approximately normal distributions with mean $μ$ and standard deviation $σ$ have:

* About 68% of data values lie within 1 standard deviation of the mean
* About 95% of data values lie within 2 standard deviations of the mean
* About 99.7% of data values lie within 3 standard deviations of the mean

**Distribution of SAT Scores**

SAT math scores were collected from a random sample of 1000 students. The dotplot (on the left) and frequency histogram (on the right) summarize the data.

|  |  |
| --- | --- |
|  |  |

These data are approximately normal with a mean of 500 and a standard deviation of 121. Use the Empirical Rule to answer the following questions.

1. Determine the interval, centered about the mean, which contains: (a) about 68% of math SAT scores, (b) about 95% of math SAT scores, and (c) about 99.7% of math SAT scores.
2. Estimate the percent of math SAT scores: (a) between 258 and 621, (b) between 500 and 863, and (c) less than 379.
3. Find the math SAT score that is: (a) greater than approximately 16% of math SAT scores, and (b) less than approximately 2.5% of math SAT scores.

The Empirical Rule provides us a way to characterize data values in a data set as unusual. Any data value that is outside the middle 95% of data values is considered *unusual.*

1. Which SAT Math scores are unusual?

**Z-scores (Standard Scores)**

It is often necessary to identify a data value’s relative position within its distribution. When the mean and standard deviation of a distribution are known, you can find a value’s relative position by calculating its *standardized score* or *z-score.* A *z*-score describes the number of standard deviations that a data value, *x*, is from the mean. The *z*-score formula is shown below.

$$z=\frac{value-mean}{standard deviation}=\frac{x-μ}{σ}$$

Every data value in a distribution has a *z*-score. *Z*-scores can be negative, positive, or zero. When a *z*-score is less than -2 or greater than 2, we say that its corresponding data value is unusual.

1. The table below displays SAT Math scores for three high school students. These three scores are from the random sample of SAT Math scores with mean 500 and sample standard deviation 121. Find the *z*-score of each student’s score.

|  |  |
| --- | --- |
| **Student** | **SAT Math Score** |
| Marcus | 630 |
| William | 450 |
| Stephanie | 750 |

1. Are any of the scores above unusual? Explain.

The previous problem shows how *z*-scores can be used to compare values from the same distribution. They are equally useful for comparing values from different distributions. Often we are told we can’t compare individuals from two different populations, but via z-scores, we can!

**Comparing Values from Different Distributions**

1. Serena Williams, the top ranked female tennis player in the world, has one of the best serves in women’s tennis. Serena’s typical first-serve speed is 120 miles per hour. Suppose first-serve speeds of female professional tennis players are bell-shaped with a mean of 110 miles per hour and a standard deviation of 5 miles per hour.

Justin Verlander, the 2011 American League Most Valuable Player (MVP) and ace pitcher for the Detroit Tigers, is an elite pitcher in Major League Baseball (MLB). Verlander’s typical fastball speed is 95 miles per hour. Suppose fastball speeds of MLB pitchers are bell-shaped with a mean of 90 miles per hour and a standard deviation of 2 miles per hour.

What is more impressive: Serena William’s s serve or Justin Verlander’s fastball?

1. The dotplots below display Connecticut and Nevada annual precipitation totals (in inches) from 1890 to 2013. Annual precipitation totals were measured from a single location in each state. The mean and standard deviation of each data set are provided below.

|  |  |
| --- | --- |
|  | Mean: $μ=43.7$ inchesStandard deviation:$σ=7.0$ inches |

|  |  |
| --- | --- |
|  | Mean: $μ=12.3$ inchesStandard deviation:$σ=3.8$ inches |

What is more likely: observing a year in Connecticut, between 1890 and 2013, in which the total precipitation is 55 inches or more, or observing a year in Nevada, between 1890 and 2013, in which the total precipitation is 8 inches or less?