AP Statistics

Chapter 2: Modeling Distributions of Data

Day 2

HW: p. 107-108, # 19, 21, 23, 31, 32

---

p. 105-107, #1, 5, 7, 9, 11, 15

1.
(a) Putting the data in order we get: 13 13 13 15 19 22 23 23 24 26 26 30 31 34 38 49 50 50 51 57. The girl with 22 pairs of shoes is the 6th smallest. Therefore, her percentile is $\frac{5}{20} = 0.25$. In other words, 25% of the girls had fewer pairs of shoes than she did.

(b) Putting the data in order we get: 4 5 5 5 6 7 7 7 7 8 10 10 10 10 11 12 14 22 35 38. The boy with 22 pairs has more shoes than 17 people. Therefore, his percentile is $\frac{17}{20} = 0.85$. In other words, 85% of boys had fewer pairs of shoes than he did.

(c) The boy is more unusual because only 15% of the boys have as many or more than he has, while the girl has a value that is more centered in the distribution. 25% have fewer and 75% have as many or more.

5. The girl in question weighs more than 48% of girls her age, but is taller than 78% of the girls her age. Since she is taller than 78% of girls, but only weighs more than 48% of girls, she is probably fairly skinny.

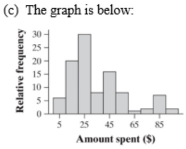
7.
(a) The highlighted student sent about 212 text messages in the 2-day period which placed her at about the 80th percentile.

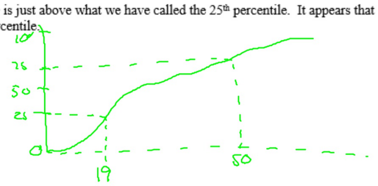
(b) The median number of texts is the same as the 50th percentile. Locate 50% on the y-axis, read over to the points and then find the relevant place on the x-axis. The median is approximately 115 text messages.

9.
(a) First find the quartiles. The first quartile is the 25th percentile. Find 25 on the y-axis, read over to the line and then down to the x-axis to get about \$19. The 3rd quartile is the 75th percentile. Find 75 on the y-axis, read over to the line and then down to the x-axis to get about \$50. So the interquartile range is $\$50 - \$19 = \$31$.

(b) The person who spent \$19.50 is just above what we have called the 25th percentile. It appears that \$19.50 is at about the 26th percentile.

(c) The graph is below:





---

11. Eleanor's standardized score, $z = \frac{680-500}{100} = 1.8$, is higher than Gerald's standardized score, $z = \frac{27-18}{6} = 1.5$.

15.
(a) Since 22 salaries were less than Lidge's salary, his salary is at the $\frac{22}{29} = 75.86$ percentile.

(b) $z = \frac{6,350,000 - 3,388,617}{3,767,484} = 0.79$. Lidge's salary was 0.79 standard deviations above the mean salary of \$3,388,617.

The Effect of Changing Units on Statistics

Sometimes, researchers change units (minutes to hours, feet to meters, etc.). Here is how measures of variability are affected when we change units.

If you add a constant to every value, the distance between values does not change. As a result, all of the measures of variability (range, interquartile range, standard deviation, and variance) remain the same. +4

Ex.: 2, 3, 4, 5, 6, 7, 8 — 6, 7, 8, 9, 10, 11, 12

	2, 3, 4, 5, 6, 7, 8	6, 7, 8, 9, 10, 11, 12
mean:	5	9
median:	5	9
range:	6	6
IQR:	4	4
Stand Dev:	2.16	2.16
Variance:	4.67	4.67

The Effect of Changing Units on Statistics

On the other hand, suppose you multiply every value by a constant.

This has the effect of multiplying the range, interquartile range (IQR), and standard deviation by that constant.

It has an even greater effect on the variance. It multiplies the variance by the square of the constant. ×2

Ex.: 2, 3, 4, 5, 6, 7, 8 — 4, 6, 8, 10, 12, 14, 16

	2, 3, 4, 5, 6, 7, 8	4, 6, 8, 10, 12, 14, 16
mean:	5	10
median:	5	10
Range:	6	12
IQR:	4	8
Stand:	2.16	4.32
Variance:	4.67	18.66

Example

The weights (in pounds) of 10 jockeys that are going to race at Arlington Park are 113, 117, 112, 113.5, 115.8, 114, 114.6, 113.5, 112.4, and 113.

Summary statistics are as follows:

Mean = 113.88 Standard deviation = 1.54
Median = 113.5 IQR = 1.6

$\frac{2.2}{113.88}$ ~~1~~ ~~x~~

$\frac{2.2x}{2.2}=\frac{113.88}{2.2}$

If the jockey's weights were converted from pounds to kilograms (2.2 pounds = 1 kilogram), what would the new summary measures be?

mean = 51.76 Stand Dev = .7
median = 51.59 IQR = .73

Example

Each jockey's clothes and saddle add about 4 pounds to the weight placed on the horse. What are the summary statistics for the weight of the 10 jockeys plus their equipment?

Summary statistics are as follows:

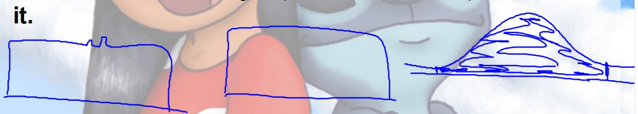
Mean = 113.88 Standard deviation = 1.54
Median = 113.5 IQR = 1.6

Mean = 117.88
Median = 117.5
Stand Dev = 1.54
IQR = 1.6

Density Curve

A density curve is a curve that describes the overall pattern of the data.

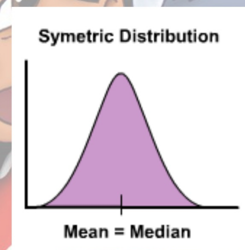
- Can be any shape (uniform, unimodal), skewed or symmetric.
- Is always on or above the horizontal axis.
- Has an area of exactly 1 (100% of the data) underneath it.



Density Curve

A symmetric curve balances at its center because the two sides are identical.

The mean and median of a symmetric density curve are equal.

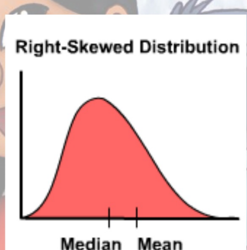


Density Curve

A density curve can be skewed to the right.

The mean is greater than the median.

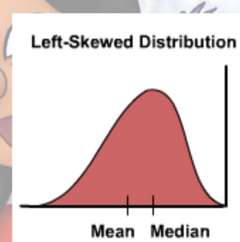
3	1
4	22
5	43216
6	7889
7	



Density Curve

A density curve can be skewed to the left.

The median is greater than the mean.

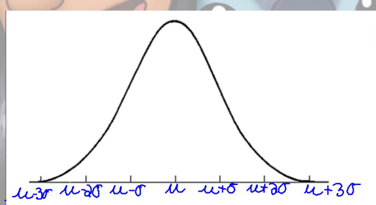


Density Curves with Normal Distributions

All normal curves have the same overall shape:

- Symmetric, single-peaked, bell-shaped.

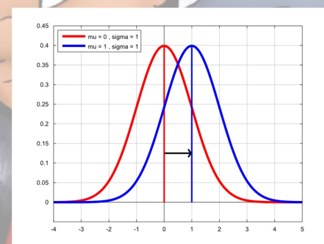
Any normal curve is described by giving its mean μ mu and the standard deviation σ. sigma



Density Curves with Normal Distributions

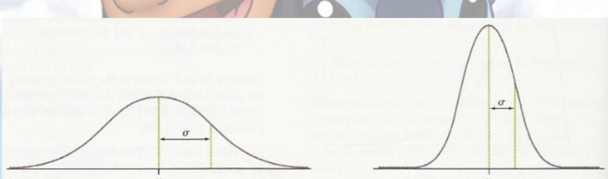
- The mean is located at the center of the symmetric curve and is the same as the median.
- Changing μ without changing σ moves the Normal curve along the horizontal axis without changing its spread.

$N(0,1)$
$N(1,1)$



Density Curves with Normal Distributions

- The standard deviation σ controls the spread of the normal curve. Curves with larger standard deviations are more spread out.
- We abbreviate the normal distribution with mean μ and standard deviation σ as $\boxed{N(\mu, \sigma)}$.



Normal Distributions and Standard Deviation

- When is the standard deviation 0?

  All the data points are the same

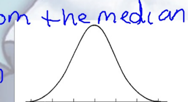
- What type of distribution gives you a small standard deviation?

  when the data points are close to the mean.

- How can you tell if a distribution is skewed vs. roughly symmetrical?

  Skewed: when the mean deviates from the median

  Symmetric: when the mean and median are about the same.



Normal Distributions

Why are the normal distributions important in statistics?

- **Many statistical inference procedures are based on normal distributions.**
  - **Scores on tests taken by many people (such as SAT exams or IQ tests)**
  - **Repeated, careful measurements of the same quantity (like the diameter of a tennis ball)**
  - **The number of heads in many tosses of a fair coin**