Edhesive AP Statistics **Unit 1 – Solutions**

**Multiple Choice:** Choose the best answer choice for the following problems.

1. Martina’s class conducted an experiment using a large number of seeds, recording the time in days that it took each seed to sprout after planting. Some of the seeds were planted in an indoor planter that was kept in the classroom, and some were planted outside. These data were used to generate the boxplots shown below.



Which of the following statements is true?

* 1. The mean time to sprout for seeds that were planted inside is approximately 20 days.
  2. The first quartile of time to sprout for seeds that were planted outside is approximately 15 days.
  3. The range in time to sprout for seeds that were planted inside is approximately 40 days.
  4. Approximately half of the seeds planted inside took longer to sprout than any seed that was planted outside.
  5. The interquartile range of the time to sprout for seeds that were planted inside is approximately 10 days.

According to the boxplots, the maximum time to sprout for seeds planted outside was 20 days, and the median time to sprout for seeds planted inside was 20 days. Since approximately half of any dataset is larger than its median, approximately half of the seeds planted inside took longer than 20 days to sprout.

1. A call center used the median instead of the mean when it advertised the amount of time that its representatives spend on the average phone call. When the call center’s data on phone call time is displayed in a histogram, it can be seen that the distribution of call times is very right-skewed. Which of the following explains why the call center would report the median instead of the mean in this situation?
   1. The mean is much smaller than the median when the data is very right-skewed.
   2. The median is exactly halfway between the minimum call time and the maximum call time.
   3. The median is not affected by the skewness of data, while the mean can be strongly affected.
   4. The median is always the preferred statistic when the underlying data is a measure of time.
   5. The mean cannot be calculated for skewed data.

Right-skewed data has a distribution with a long tail on the right, and these values that are much larger than normal can inflate the mean of the dataset so that it does not very well represent the “average” data point. The median, in contrast, is not affected by such skewness or by outliers.

1. The music teachers at a high school surveyed students in the marching band, asking them how long they have been playing their instrument. The results are reported by the students’ instrument type in the table below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | <1 year | 1 to 3 years | 3 to 5 years | >5 years | Total |
| Brass | 4 | 15 | 10 | 2 | 31 |
| Percussion | 6 | 12 | 3 | 3 | 24 |
| Woodwind | 3 | 11 | 14 | 4 | 32 |
| Total | 13 | 38 | 27 | 9 | 87 |

The observed number of woodwind players who have played their instrument for less than 1 year was 3. What would be the expected number of woodwind players in this sample who have played their instrument for less than 1 year, if the length of time playing an instrument were independent from instrument type?



If A and B are independent, then *P*(A and B) = *P*(A) · *P*(B). So, under the assumption that length of time playing an instrument is independent from instrument type, one would calculate *P*([woodwind] and [<1 year]) = *P*([woodwind]) · *P*([<1 year]). From the collected data, *P*([woodwind]) ≈ and *P*([<1 year]) ≈ , so that the probability of a student both playing a woodwind instrument and having played it for less than one year is ≈ 0.055.

In the current sample, 87 students were surveyed, so the expected *number* of students with this characteristic = =

1. The Weekender Movie Theatre randomly sampled 25 of its movie showings, recording the number of ticket purchases for each showing. The histogram below displays the data collected by the movie theatre.



In which interval is the median number of ticket purchases located?

* 1. 20 tickets to less than 40 tickets
  2. 40 tickets to less than 60 tickets
  3. 60 tickets to less than 80 tickets
  4. 80 tickets to less than 100 tickets
  5. 100 tickets to less than 120 tickets

There were 10 movie showings that had ticket sales of less than 40, and 12 showings that had ticket sales of 60 or above. The remaining 3 showings had ticket sales in the 40 to 60 interval, and the median is the largest of these three.

1. A local charity called 100 of their previous supporters at home in December to solicit end-of-year donations. A manager at the charity was curious if there was a relationship between whether the supporters answered the phone and whether they made an end-of-year donation. The data collected by the manager is displayed in the table below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Made an End-of-Year Donation | |  |
|  |  | Yes | No | Total |
| Answered  Phone Call | Yes | 30 | 10 | 40 |
| No | 25 | 35 | 60 |
|  | Total | 55 | 45 | 100 |

Which of the following statements is true for these 100 supporters?

* 1. They were more likely not to answer the phone call than to answer it.
  2. They were more likely not to make a donation than to make one.
  3. There were more supporters who answered the phone call and made a donation than supporters who did not answer the phone call and did not make a donation.
  4. Answering the phone call but not making a donation was more likely than not answering the phone call but making a donation.
  5. Not answering the phone call but making a donation was more likely than not answering the phone call and not making a donation.

According to the table, there were a total of 40 supporters who answered the phone call and 60 supporters who did not answer the phone call. Thus, not answering the phone call was the more likely outcome.

1. A graph (not shown) of the distances achieved in the qualifying round of the 2012 Olympic men's long jump reveals that the distribution is skewed to the left. Which of the following statements is the most reasonable conclusion that can be drawn from this fact about the long jump distances?
   1. The mean is greater than the median.
   2. The median is the average of the first quartile and the third quartile.
   3. There were fewer jump distances between the first quartile and the median than there were between the median and the third quartile.
   4. There were more jump distances below the mean than there were jump distances above the mean.
   5. The difference between the minimum and the first quartile is greater than the difference between the third quartile and the maximum.

Left-skewed data has a distribution with a long tail on the left. The first quarter of the data includes this long tail, so the first quartile is relatively far from the minimum. In contrast, between the third quartile and the maximum there is no long tail and so the difference is comparatively less.

1. A team of biologists observed 100 duck nests and recorded the clutch size (number of eggs) for each nest at the start of the nesting season. The histogram below shows the distribution of clutch sizes for these 100 nests.



Duck Nests

Which of the following values is closest to the standard deviation for this data set?

* 1. 0.5 eggs
  2. 2 eggs
  3. 5 eggs
  4. 9 eggs
  5. 13 eggs

From the histogram, it can be estimated that the mean of the data set is approximately 9 and that approximately two-thirds of the data falls in the range from 7 to 11. Thus, 2 eggs is a reasonable estimate for the standard deviation.

1. Aruna works as a saleswoman and earns commission pay, with mean earnings of $175 in commission pay per day. She had an unusually strong sales day on Monday, earning $285 in commission. If she works the rest of the week (Tuesday through Friday) and averages the typical $175 per day over those four days, what will be her average daily commission earnings for this week?
   1. $140
   2. $175
   3. $197
   4. $230
   5. $246

If Aruna averages $175 per day over the remaining four days in the week, her total earnings for those days will be 4 · $175 = $700. Added to her earnings from Monday, her total earnings over the whole week will be $285 + 4 · $175 = $985. This total is taken over 5 days, so the average is

1. A factory manager conducted a review of 10 machines in his factory, recording for each machine the number of defective parts that were produced in a week. The data collected by the manager is summarized below:

|  |  |
| --- | --- |
| Min = 1  Lower quartile = 7  Median = 9 | Max = 13  Upper quartile = 10  n = 10 |

Which of the following statements is true?

* 1. The smallest observation is 1 and it is an outlier. No other observations in the data set could be outliers.
  2. The largest observation is 13 and it is an outlier. No other observations in the data set could be outliers.
  3. 1 is an outlier and it is possible that there are other outliers at the low end of the data set. There are no outliers at the high end of the data set.
  4. 13 is an outlier and it is possible that there are other outliers at the high end of the data set. There are no outliers at the low end of the data set.
  5. Neither 1 nor 13 is an outlier. There are no outliers in the data set.

An outlier is a data point that lies outside the upper or lower fence which is 1.5 times the IQR below Q1 or above Q3. For this data set, the IQR = 10 – 7 = 3. The lower fence is at 7 – 1.5(3) = 2.5. The upper fence is at 10 + 1.5(3) = 14.5. The minimum observation is an outlier below the lower fence. Because the statistics do not indicate the next highest point from the minimum, it is possible that there are more than one outlier between 1 and 2.5.

1. A bus company examined the number of passengers who rode a particular interstate route for the past year. Throughout the year, there were 450 trips made on this route, with about half running on a weekday (Monday through Thursday) and the other half running on a weekend (Friday, Saturday, or Sunday). The figure below summarizes the numbers of passengers carried on these 450 trips.



Which of the following statements about the medians and interquartile ranges (IQRs) of the number of passengers for weekday and weekend trips is true?

* 1. The median for weekday trips is greater than the median for weekend trips, and the IQR is also greater for weekday trips.
  2. The median for weekday trips is greater than the median for weekend trips, and the IQR is less for weekday trips.
  3. The median for weekday trips is less than the median for weekend trips, and the IQR is also less for weekday trips.
  4. The median for weekday trips is less than the median for weekend trips, and the IQR is greater for weekday trips.
  5. The medians for weekday and weekend trips are different, but the IQRs are the same.

According to the boxplot, the median for weekday trips is approximately 15 passengers while for weekend trips it is approximately 25, and the IQR for weekday trips is approximately 15 passengers while for weekend trips it is approximately 10.

1. Of the four data sets summarized in the histograms below, which has the smallest proportion of data below its mean?
   1. Data set 1
   2. Data set 2
   3. Data set 3
   4. Data set 4
   5. The histograms do not provide sufficient information to determine which data set has the smallest proportion.

Because Data set 4 is strongly left-skewed, its mean is less than its median, so that less than 50% of the data is below the mean. In contrast, Data sets 1 and 2 are roughly symmetric, meaning that about 50% of the data is below the mean for each of them; Data set 3 is right-skewed, meaning that more than 50% of the data is below the mean.

1. The following summary statistics are based on systolic blood pressure measurements for 300 athletes. Blood pressure is measured in millimeters of mercury (mmHg).

Mean: 130 mmHg

Median: 121 mmHg

Standard Deviation: 8 mmHg

First Quartile: 114 mmHg

Third Quartile: 133 mmHg

Number of Observations: 300

About 150 of the athletes in this sample had systolic blood pressure that was:

* 1. Less than 114 mmHg
  2. Between 114 and 133 mmHg
  3. Less than 133 mmHg
  4. More than 130 mmHg
  5. Between 122 and 138 mmHg

For any data set, about 50% of the data falls between the first and third quartiles.

1. The stemplot below shows individual student grades for two college classes, Class X and Class Y, who took the same mass exam.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Class X |  | | Class Y | |  | |
| 0, 0, 0 | 10 |  | |  | |
| 8, 8, 3, 3, 3, 3, 1 | 9 | 1, 1, 2, 4 | |  | |
| 9, 9, 5, 4, 4, 4, 3, 3, 1 | 8 | 2, 2, 2, 3, 4, 4, 5, 7, 7, 8, 9 | |  | |
| 7, 6, 2, 1 | 7 | 3, 3, 3, 4, 4, 7, 7, 9, 9 | |  | |
|  | 6 | 2, 4, 8 | |  | |
|  |  | |  | | 7 | 5 = 75 | |

Which of the following statements is true?

* 1. The first quartile for Class X was less than the first quartile for Class Y.
  2. The median for Class X was less than the median for Class Y.
  3. The mean for Class X was higher than the mean for Class Y.
  4. The range for Class X was larger than the range for Class Y.
  5. The interquartile range for Class X was larger than the interquartile range for Class Y.

The data for Class X and Class Y can be summarized as follows:

|  |  |
| --- | --- |
| Class X  Min = 71  Q1 = 83  Median = 89  Mean = 87.7  Q3 = 93  Max = 100 | Class Y  Min = 62  Q1 = 74  Median = 82  Mean = 80.5  Q3 = 87  Max = 94 |

1. The table below shows the sample size, the mean, and the standard deviation for two samples of data.

|  |  |  |  |
| --- | --- | --- | --- |
|  | *n* | Mean | Std. Dev. |
| Sample A | 38 | 4.46 | 1.31 |
| Sample B | 43 | 5.26 | 1.02 |

What is the mean for the combined sample of 81 observations?

(A)

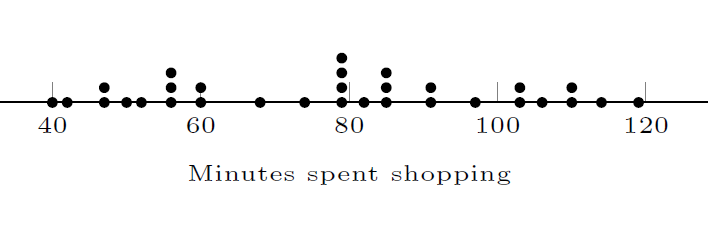
(B)

(C)

(D)

(E)

The sum of all observations in Sample A is 38(4.46) and the sum of all observations in Sample B is 43(5.26), so the sum of all observations in the combined sample is 38(4.46) + 43(5.26). Since the combined sample has 81 observations, its mean is therefore

1. The dotplot below displays the total number of minutes that a sample of 31 people spent grocery shopping in a single week.

Which of the following represents the shopping time at the 25th percentile?

1. 50 minutes
2. 55 minutes
3. 75 minutes
4. 80 minutes
5. 100 minutes

The data point at the 25th percentile is quartile 1. With 31 data points, Q1 is between the 8th data point.

1. A distribution of start-up times for computers is skewed left. Which of the following is the best estimate for the z-score of the first quartile?
2. 0.25
3. -0.67
4. 0.598
5. -0.25
6. A z-score cannot be approximated with the information given.

A standardized score, z-score, can only be estimated with a normal distribution.

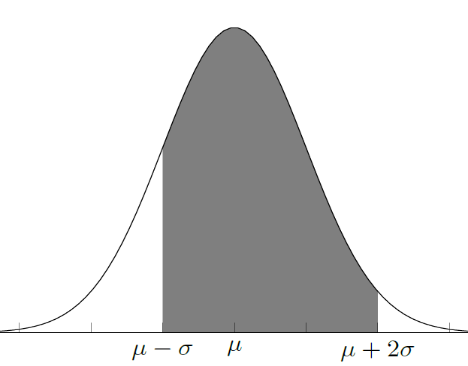
1. A data set of test scores is transformed by applying the following rule:

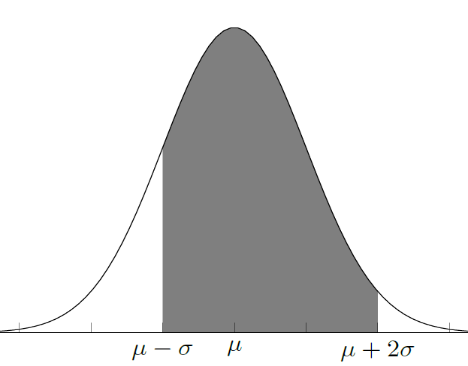
*Transformed test score = 90 + 2.2(raw test score)*

Which of the following statements is true?

1. The IQR of the transformed data equals 90 + 2.2(the IQR of the raw test score).
2. The range of the transformed data equals 90 + 2.2(the range of the raw test score).
3. The median of the transformed data equals 90 + 2.2(the median of the transformed test score).
4. The standard deviation of the transformed data equals the standard deviation of the raw test score.
5. The mean of the transformed data equals 2.2(the mean of the raw test score).

Center measurements are translated by multiplication and addition while spread measurements of a distribution are translated by multiplication only.

1. For a certain brand of climbing rope, the maximum force (measured in kiloNutens) which can be applied before the rope breaks is approximately normally distributed with a mean of 10kN and a standard deviation of 0.2kN. The shaded figure below represents which of the following probabilities?



9.8

10.4

10

10.2

10.6

9.6

9.4

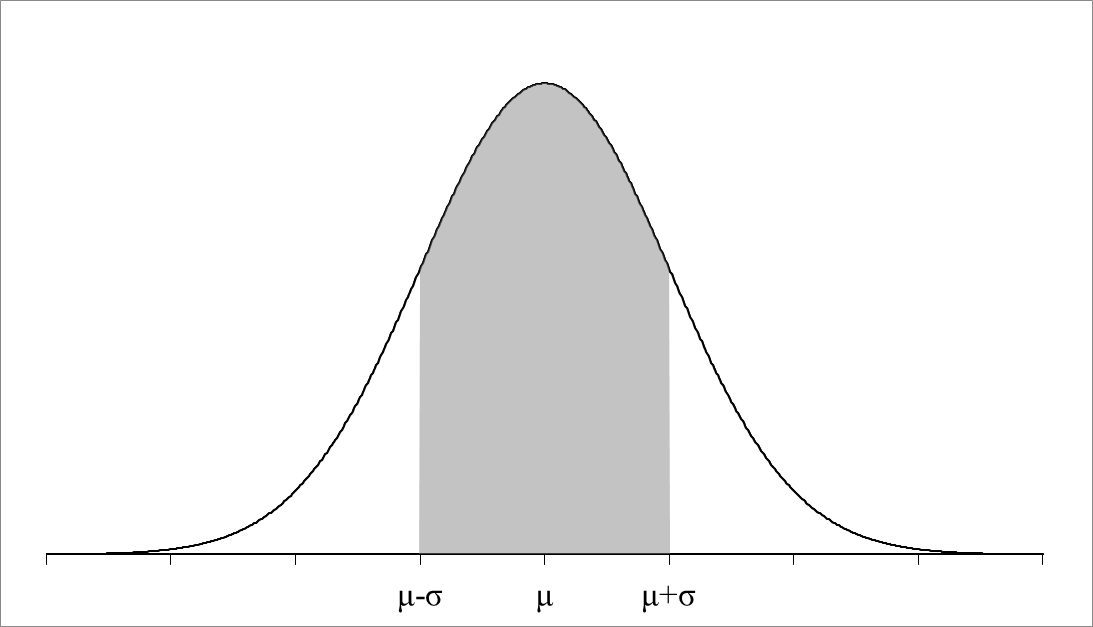
1. The probability that the maximum force applied before breakage of a randomly selected rope of this brand is between 9.8 kN and 10.4 kN.
2. The probability that the maximum force applied before breakage of a randomly selected rope of this brand is between 9.6 kN and 10.0 kN.
3. The probability that the maximum force applied before breakage of a randomly selected rope of this brand is between 9.6 kN and 10.4 kN.
4. The probability that the maximum force applied before breakage of a randomly selected rope of this brand is between 10.0 kN and 10.4 kN.
5. The probability that the maximum force applied before breakage of a randomly selected rope of this brand is between 9.8 kN and 10.6 kN.
6. Weight, in grams, is measured for each person in a sample. After the data are collected, all the weight measurements are converted from grams to kilograms by dividing each measurement by 1000. Which of the following statistics will remain the same for both units of measure?
7. The minimum of the weight measurements.
8. The interquartile range of the weight measurements.
9. The mean of the weight measurements.
10. The z-scores of the weight measurements.
11. The standard deviation of the weight measurements.

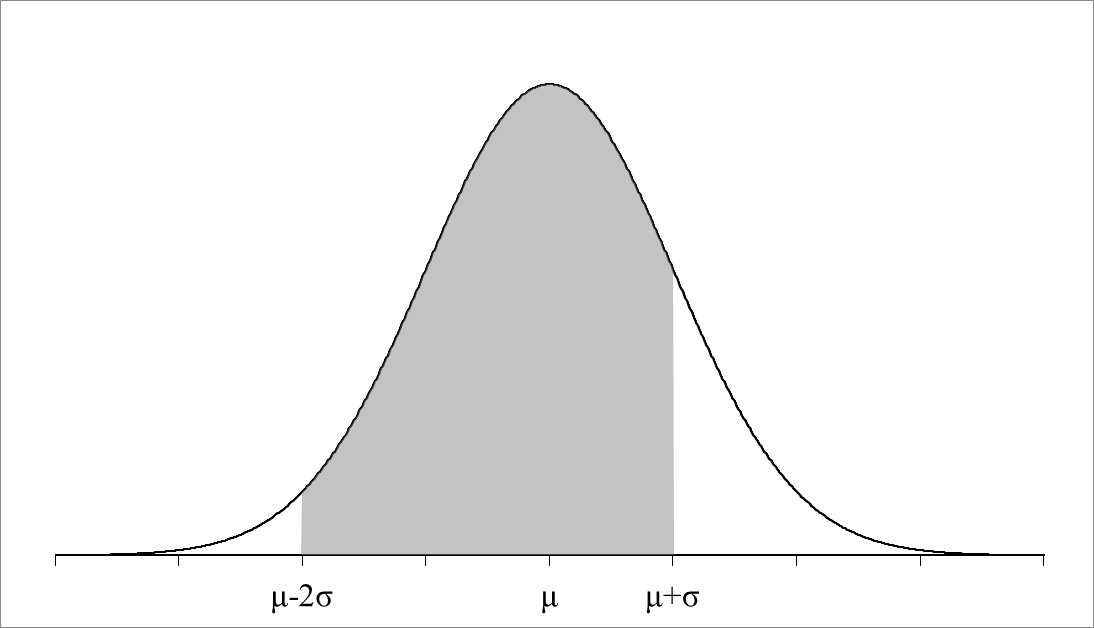
Standardized score (z-score) are resistant to unit changes.

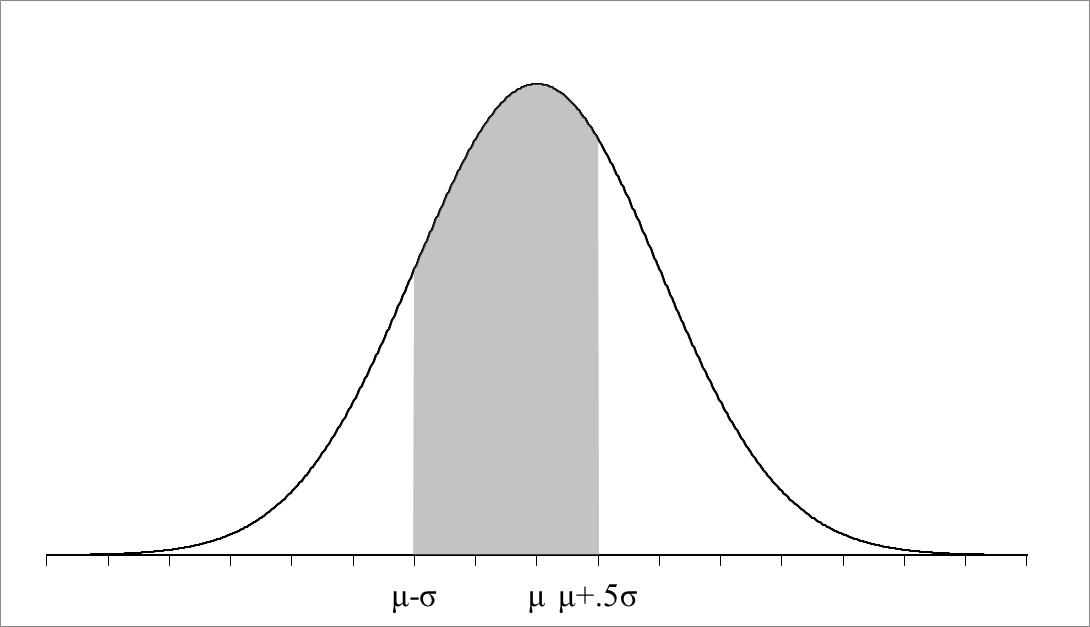
1. A sandwich shop near a university sells subs to students and faculty. The owner of the shop believes the length of the subs he sells are normally distributed with a mean of 25 cm. If the owner’s beliefs are correct, of the intervals below, which of the following will contain the largest proportion of the subs in the distribution of length?
   1. 25 cm to 29 cm
   2. 21 cm to 25 cm
   3. 18 cm to 22 cm
   4. 23 cm to 27 cm
   5. 15 cm to 19 cm

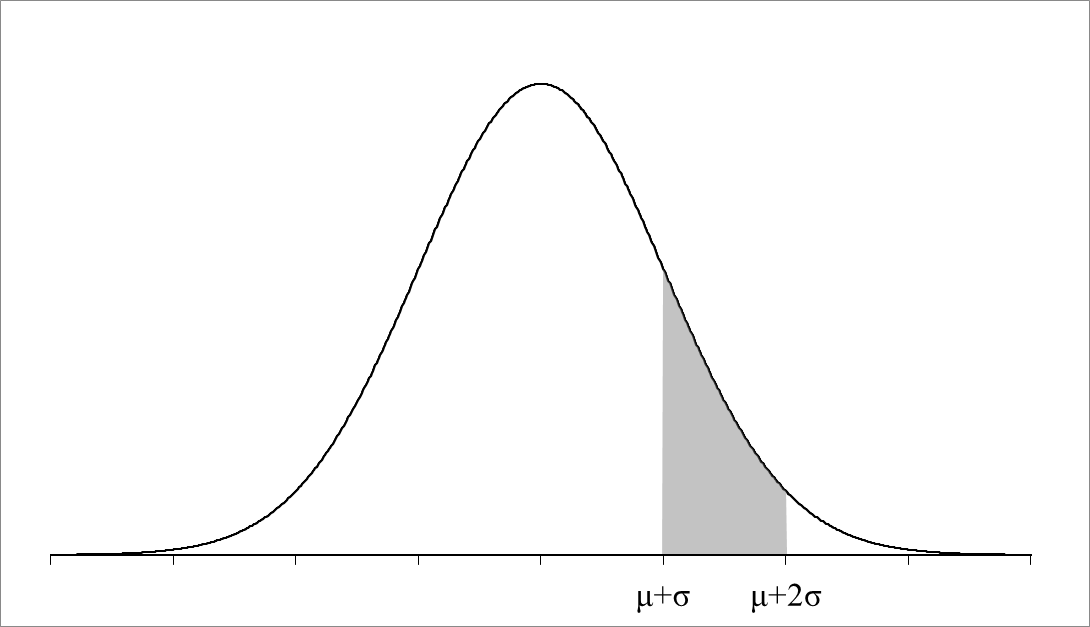
The lower limit in each of the options is always 4 less than the upper limit, only choice d is centered at the mean, which implies a larger proportion relative to the other options.

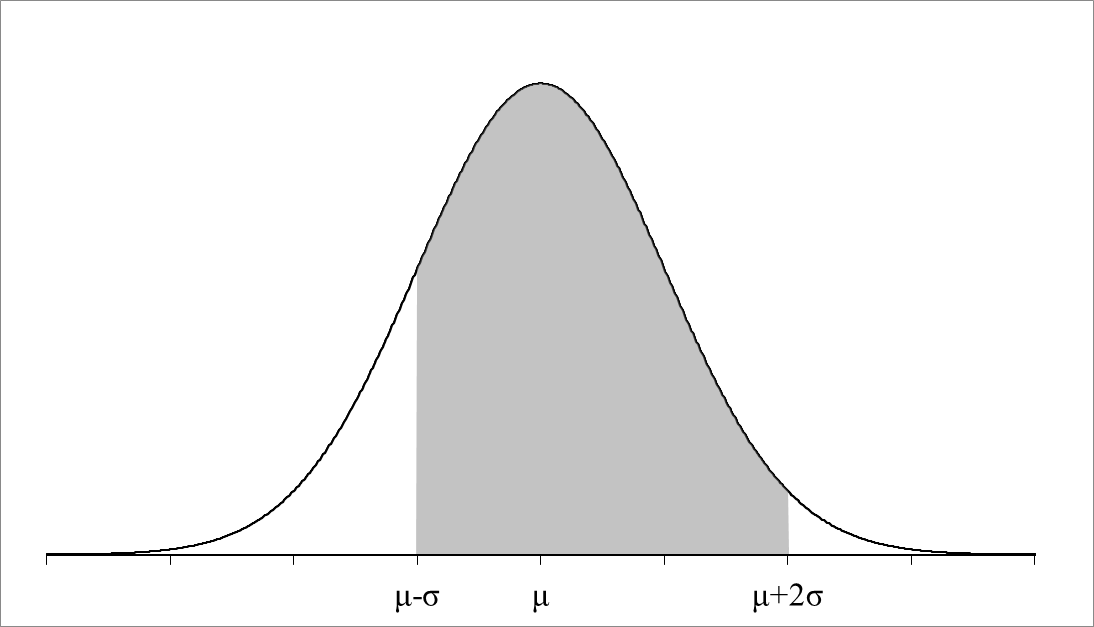
1. The amount of bamboo eaten by giant pandas per day is approximately normally distributed with a mean of 30 pounds and a standard deviation of 5 pounds. Which of following shaded regions best represents the probability that a randomly selected panda will eat between 20 pounds and 35 pounds of bamboo on a given day?

(A)

(B)

(C)

(D)

(E)

The shaded region best corresponding to the probability that a randomly selected panda will consume between 20 and 35 pounds is found by first determining how far the lower and upper z-score

lower limit = -2 upper limit

1. The distribution of the weight of adult housecats is approximately normally distributed with a mean of 8.9 pounds and a standard deviation of 1.32 pounds. Lawrence is an adult house cat and his weight is 1.56 standard deviations above the mean. Felix, a neighbouring housecat, is at the 90th percentile of the weight distribution. At what percentile is Lawrence’s weight, and how does his weight compare to Felix’s weight?
   1. Lawrence is in the 6th percentile of the weight distribution and he is heavier than Felix
   2. Lawrence is in the 6th percentile of the weight distribution and he is lighter than Felix
   3. Lawrence is in the 94th percentile of the weight distribution and he is heavier than Felix
   4. Lawrence is in the 94th percentile of the weight distribution and he is lighter than Felix
   5. Lawrence is in the 6th percentile of the weight distribution and he is lighter than Felix

Lawrence’s weight is 1.56 standard deviations above the mean. Using a z-table or calculator, we find that Lawrence is heavier than 94.06 percent of the population. Given that Felix is only heavier than 90 percent of the population of housecats, it follows that Lawrence is heavier.

1. The distribution of the weight of adult housecats is approximately normally distributed with a mean of 8.9 pounds and a standard deviation of 1.32 pounds. How does the weight of a housecat whose weight is at the 98th percentile compare with the mean weight of adult house cats?
   1. 2.711pounds greater than the mean
   2. 2.053 pounds greater than the mean
   3. 1.293 pounds greater than the mean
   4. 2.053 pounds less than the mean
   5. 2.711 pounds less than the mean.

Use a table or technology to find the z-score for 98% of a normal curve. z = 2.0537

x – 8.9 = difference in means

11.611 – 8.9 = 2.711 lbs.

2.711 = x – 8.9

x = 11.611

1. The weight of adult men is approximately normally distributed with a mean weight of 195.5 pounds and a standard deviation of 29 pounds. The weight of adult women is approximately normally distributed with a mean weight of 166.2 and a standard deviation of 30. A certain adult woman weighs 135 pounds. This woman would have roughly the same standardized weight as an adult male who weighed what?
   1. 165.34 pounds
   2. 225.66 pounds
   3. 195.5 pounds
   4. 175 pounds
   5. 107.72 pounds

Find the standardized score for the woman.

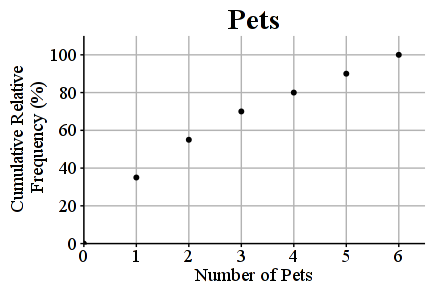
, where and

To find the weight of an adult male who would have the approximately the same z-score as the women

, where is the man’s weight,

x = 165.34 pounds.

1. An AP Statistics class is surveyed for the number of pets each student owns. The cumulative frequency distribution for the gathered data is shown below. Which of the following is the percentage of students who own 3 or more pets?



* 1. 20
  2. 30
  3. 50
  4. 60
  5. 70

The number of students with 3 or more pets is at the 70th percentile, so 30% students responded with 3 or more pets on the survey.

1. Consider *n* test scoresThe mean and standard deviation of the scores is =5.5 and =3.6, respectively. Each score is transformed according to *y*=0.56*x* -1.76, where *y* is the transformed score and *x* is the original score. Which of the following is the mean of the transformed scores?
   1. 1.9
   2. 1.32
   3. 0.292
   4. 3.08
   5. 3.54

The test scores are first multiplied by 0.56 then transformed 1.76 units to the left. The mean of the test scores must be transformed to find the mean of the transformed scores. That is,

*y* = 0.56(5.5)-1.76

*y* = 1.32

1. A soccer competition counts the number of goals a player makes on an empty net in one minute. The number of goals in the first round of the competition is normally distributed with a mean of 44 and a standard deviation of 16. The number of goals in the second round of the competition is predicted to be normally distributed with a mean of 58 and a standard deviation of 10. A player scores 50 goals in the first round. If the player participated in the second round, how many additional goals would improve their first round score?
   1. 1
   2. 3
   3. 6
   4. 9
   5. 12

The data is normally distributed. Using the normalization equation, the player has a *z* value of 0.375, shown below.

The results in the second round of the competition are expected to be normally distributed with a different mean and standard deviation. We are interested in how many additional goals the player must score to achieve a higher *z* value.

The player must score 62 goals, which is an additional 12 goals.

1. Which of the following statements is true when the number of erasers in a classroom is skewed to the left.
   1. The percentage of erasers within half of the standard deviation of the mean is approximately 68%.
   2. The percentage of erasers within one standard deviation of the mean is approximately 68%.
   3. The percentage of erasers within two standard deviations of the mean is approximately 95%.
   4. The percentage of erasers within three standard deviations of the mean is approximately 99.7%.
   5. The percentage of erasers within one standard deviations of the mean is cannot be calculated with the information given.

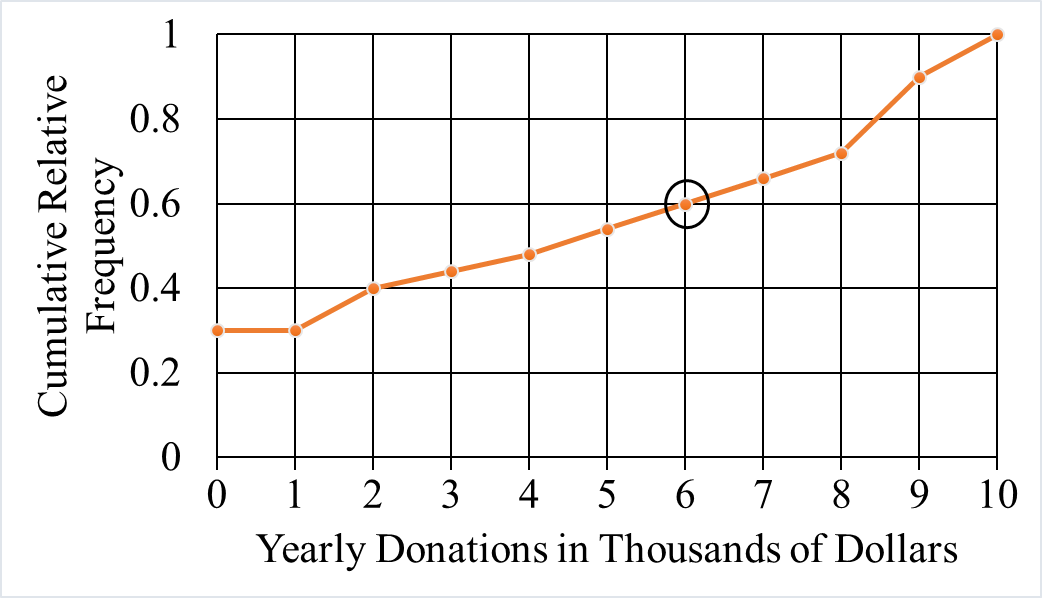
Data that is skewed does not follow a normal distribution; therefore the assumptions using the Empirical Rule are not applicable.

1. Which of the following statements is true when a median of a data set is larger than the mean?
   1. The data set is skewed to the left.
   2. The data set is skewed to the right.
   3. The data set is normally distributed.
   4. The data set is normally distributed if larger data set values are removed.
   5. The data set is normally distributed if smaller data set values are removed.

The median of a distribution is resistant to outliers, while the mean is not. If the mean is smaller than the median, it has been pulled towards lower values resulting from a left skewed distribution.

**Free Response – Solutions**

1. A large university keeps track of the yearly donations given by alumni. The figure below presents the cumulative relative frequency plot of the past year’s donations (in thousands of dollars) from the alumni.



* 1. In the context of the question, what information is conveyed by the circled point on the graph?

The circled point approximates that 60% of the alumni donated $6,000 or less.

* 1. What proportion of alumni donated between 2 and 6 thousand dollars during the past year?

The proportion of alumni who donated $6000 or less is 0.6 and $2000 or less is 0.4. The approximate proportion of alumni who donated between 2 and 6 thousand dollars is 0.6 – 0.4 = 0.2, or 20%.

* 1. Between donations of 0 and 1 thousand dollars, the cumulative frequency plot is flat. Interpret this information in the context of the question.

There are no alumni who donated between $0 and $1000.

* 1. A selection of artisan cheeses will be sent by the university to the alumni who submitted donations in the top 60 percent of donations. What is the minimum amount an alumnus would have to donate to receive the selection of artisan cheeses from the University?

The top 60% of donations would occur at the 40th percentile of the distribution, which represents $2000. Any alumni who donates at least $2000 will be in the top 60%.

1. Clownfish length is normally distributed with a mean of 13cm. The longest clownfish in the 99th percentile are 15cm.
2. What is the standard deviation of the clownfish size distribution? Justify your answer based on the procedure that you described in (a).

The *z-* score for the 99th percentile is 2.33. Use the normalization equation to solve for the standard deviation.

We are given values for x (15cm), the mean (13cm) and the *z* -score (2.33). We are able to solve for the standard deviation.

The standard deviation of the clownfish size approximately is 0.858 cm.

1. Local news reports that a pet store has a clownfish 10cm long. What percentage of clownfish is expected to be less than 10cm? Is this news story likely? Explain.

We use the normalization equation to find the corresponding *z* value. We then use Table A to find the probability of being below that *z* value.

P( z < -3.49) = 0.0002

The probability of being below z = -3.49 is 0.0002; therefore, it is possible to have a clownfish over 10 cm but not likely.

1. If the local news wants to recant the story by reporting the length of the top 5% of the fish, what would be the smallest clownfish to be in the new story?

The z-score for 95% (100% - 5%) of the curve is 1.644. Use the normalization equation to solve for the standard deviation.

x = 14.411

The smallest clownfish in the top 5% is approximately 14.41 cm long.

1. The length for commercials is normally distributed with a mean of 30 seconds and a standard deviation of 3 seconds.
2. Commercial company A makes commercials that are 28 seconds long. Determine the percentage of commercials that are shorter than 25 seconds long?

P(z < -0.667) = 0.2525

The percentage of commercials that are shorter than 25 seconds long is 25.25%.

1. With the same standard deviation, what new mean is required if the commercial company would like only 8% of commercials to be less than 25 seconds long?

Approximately 8% of the distribution is below the *z* –score of -1.4.

2

The mean length of commercials for company A is 29.2 seconds.

1. Commercial company B says 1% of their commercials are less than 29 seconds and 6% of their commercials are over 33 seconds. What is the mean and the standard deviation for commercial company B?

From Table A, 1% of the distribution is below the *z* –score of -2.3 and 94% (100%-6%) of commercials are below the *z* –score of 1.56. We use two normalization equations to solve for the mean and the standard deviation.

We have two equations with two unknowns, we can use a system of equations to solve for the mean and the standard deviation. The mean is 31.38. The standard deviation is 1.03.

1. An office manager wants to evaluate the performance of the workers in her office in order to calculate bonuses. She evaluates the workers on two traits: typing speed and speed of dealing with phone calls from clients.

To measure typing speed she uses a computer program to calculate the number of characters typed in one hour of typing for each employee. The higher number of characters typed indicate more desirable (faster) typing speeds. The mean number of characters typed is 8000, the standard deviation of characters typed is 150 and the upper quartile of characters typed is 8080, as shown in the table below.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mean | Standard Deviation | Upper Quartile |
| Characters typed in an hour | 8000 | 150 | 8080 |

* 1. Based on the relationship between the mean, standard deviation and upper quartile, is it reasonable to believe that the distribution of the characters typed in one hour is approximately normal? Explain your answer.

It is not likely that this data is approximately normal. The upper (3rd) quartile has a z-score of 0.53, whereas the z-score for the upper quartile in a normal distribution is approximately 0.67.

* 1. The office manager wants to award a bonus to either of the two employees labelled Employee A and Employee B based on their performance. The manager considers both typing speed and speed of dealing with clients on the phone to be of equal importance when making this decision. Using the information about the means and standard deviations of the measures of typing speed and speed at dealing with clients on the phone, and the results of the employees in the table below decide which employee she should award the bonus to. Justify your answer.

|  |  |  |
| --- | --- | --- |
|  | Employee A | Employee B |
| Words typed in an hour | 8320 | 8400 |
| Time taken to complete one phone call | 990 | 1080 |

|  |  |  |
| --- | --- | --- |
|  | Employee A | Employee B |
| Words typed in an hour (z-score) | 2.13 | 2.67 |
| Time taken to complete one phone call (z-score) | 0.56 | 1.56 |

Employee B has a slightly higher z-score for the words typed in an hour than employee A, which is more desirable. However Employee B has a much higher z-score than Employee A for the time taken to complete one phone call, which is much less desirable. The manager should therefore award the bonus to Employee A.

1. The manager measures how fast her workers deal with customers on the phone by recording the time taken in seconds to complete one phone call for each employee. The smaller times indicate faster (more desirable) dealings with clients. The mean time taken to complete a call is 940 seconds with a standard deviation of 130 seconds, as shown in the table below.

|  |  |  |
| --- | --- | --- |
|  | Mean | Standard Deviation |
| Time taken to complete one phone call | 940 seconds | 90 seconds |

One worker takes 685 seconds to complete a phone call. Calculate the z-score for this time and interpret the result.

The z-score for this time is -2.83. This result is therefore a fairly extreme outlier in the negative (desirable) direction. This worker therefore could be said to be extremely fast at dealing with clients on the phone and this is a very desirable trait.