

# Northview Collegiate

Unit

Name:

Course Code:

Date:



K/U:	T/I:	A:	C:
------	------	----	----

## 1. K/U (5)

### Solution

- a) This data point shows that when the local news predicted a 50% chance of precipitation, it actually rained only about 30% of the time.
- b) For most of the predictions, the national weather website was closer to perfect than the local news. The local news consistently overestimated the chance of rain, in some cases by as much as 30 percentage points. The national weather website's predictions were more accurate than the local news except where their predictions were between 0% and just over 20% PoP.
- c) Environment Canada collects weather data at many locations in small geographic areas. The national weather website may be forecasting for a larger area than the local news. If precipitation is recorded in a localized region, it may be outside the area forecast by the local news, but included in the forecast for the national weather website. Also, although the raw data come from the same source, the calculations used to make the predictions may be different. This results in different analyses of the same data.

- a) This is a simple random sample. The draw includes all possible students, so each person has an equally likely chance of being selected.
- b) Since the employees are ordered, and every tenth person is chosen, this is a systematic sample. As long as the first person is chosen randomly, everyone has an equal chance of being selected.
- c) If 25% of each town is sampled, this is a stratified sample since the number from each town in the sample would be in the same proportion as in the population (the company would survey 4764 people from Huntsville, 210 from Kearney, 579 from Emsdale, and 195 from McMurrich). Everyone has an equal opportunity to be selected since the 25% of people being sampled are randomly selected from each town.

- d) Since the residents choose whether or not they respond to the survey, it is a voluntary response survey. Even though all residents received the survey, the data will reflect responses only from those who are interested. Usually only people who feel strongly one way or another will respond. Thus, the entire population will not be represented in the sample.
- e) Since the reporter surveys anyone who is walking on the street, the data are collected in a way that is easily accessible. So, this is a convenience sample. The entire population is not represented since only people in that area at that time are surveyed.
- f) This is a cluster sample because the population is divided into groups, a number of sections are randomly chosen, and all farms in those sections are surveyed. Since there is random selection in the first set of groups and sections, all groups and sections are equally likely to be chosen. It is more efficient for the researchers to visit the farms within five areas than it would be to travel to different farms throughout the whole province.
- g) This population has a hierarchy of organization and random sampling occurs at each level of the organization, so this is multistage sampling. Each student is equally likely to be chosen since there is random sampling at every level.

a) Calculate the weighted mean:

$$\begin{aligned}\mu_w &= \frac{90(0.25) + 79(0.20) + 82(0.15) + 70(0.10) + 85(0.15)}{0.85} \\ &= \frac{70.35}{0.85} \\ &\approx 82.76\end{aligned}$$

b) Calculate the unweighted mean:

$$\begin{aligned}\mu &= \frac{90 + 79 + 82 + 70 + 85}{5} \\ &= \frac{406}{5} \\ &= 81.2\end{aligned}$$

Without the weighted mean, Kyle would receive a lower average mark because the categories he excels in are worth more when calculated using the weighted mean.

c) The final exam is worth 15%. Determine the final exam score,  $E$ , needed for Kyle to receive 84% in the course.

$$\begin{aligned}84 &= \frac{90(0.25) + 79(0.20) + 82(0.15) + 70(0.10) + 85(0.15) + E(0.15)}{1.00} \\ 84 &= 70.35 + E(0.15) \\ 84 - 70.35 &= 0.15E \\ 13.65 &= 0.15E \\ 91 &= E\end{aligned}$$

Kyle must receive 91% on his final exam to finish the course with an 84%.

**K/U b)**

Order the data from least to greatest.

15 16 20 21 21 23 24 25 25 25 26 28 28  
28 28 30 30 31 32 34 36 36 37 38 40

- a) There are 25 data values. To calculate the 80th percentile, use the formula for percentile rank.

$$\begin{aligned} R &= \frac{p}{100}(n + 1) \\ &= \frac{80}{100}(25 + 1) \\ &= 20.8 \end{aligned}$$

Round down to 20. Determine the midpoint of the 20th and 21st measurements.

Why do you round down to 20?

$$\begin{aligned} \text{80th percentile} &= \frac{34 + 36}{2} \\ &= 35 \end{aligned}$$

The 80th percentile is a mark of 35. This means that 80% of the data are below 35.

- b) A mark of 25 is the 8th ranked mark. It is also the 9th and 10th marks.

There are 7 data values less than 25, so  $L = 7$ .

There are 3 data values equal to 25, so  $E = 3$ .

$$\begin{aligned} p &= 100 \frac{(L + 0.5E)}{n} \\ &= 100 \frac{(7 + 0.5(3))}{25} \\ &= 34 \end{aligned}$$

---

A mark of 25 is in the 34th percentile. This means that 34% of the data are below 25.

c) A mark of 40 is the 25th ranked mark.

There are 24 data values less than 40, so  $L = 24$ .

There is 1 data value equal to 40, so  $E = 1$ .

---

• Chapter 6

---

$$\begin{aligned} p &= 100 \frac{(L + 0.5E)}{n} \\ &= 100 \frac{(25 + 0.5(1))}{25} \\ &= \frac{100(24.5)}{25} \\ &= 98 \end{aligned}$$

A mark of 40 is in the 98th percentile.

A.

c)  $\sigma = 1.429$  years

The standard deviation would increase because age 18 is farther from the mean and the spread of the data would increase.

Population formulas were used here because all participants were included.

d)  $\sigma = 1.359$  23 years

The standard deviation would decrease because age 16 is closer to the mean and the spread of the data would decrease.

e)  $\sigma = 1.395$  48 years

Since values increase by the same amount, the spread will not change. The standard deviation would be unchanged.

f)  $\mu + \sigma = 17.86$  years

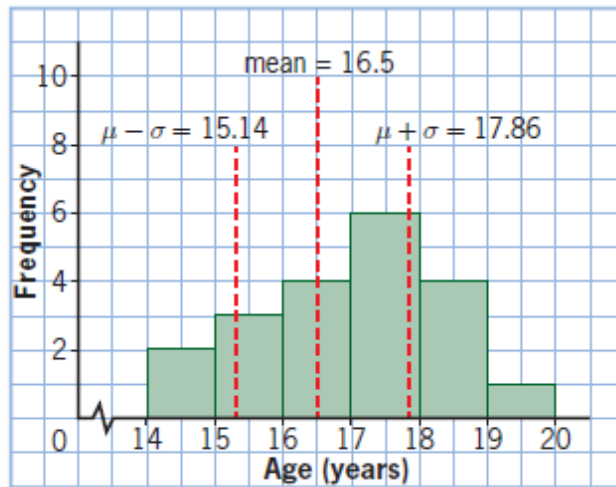
$\mu - \sigma = 15.14$  years

Ages greater than 17 or less than 16 are more than one standard deviation from the mean.

f)  $\mu + \sigma = 17.86$  years

$\mu - \sigma = 15.14$  years

Ages greater than 17 or less than 16 are more than one standard deviation from the mean.



With which type of data (ordinal, quantitative, categorical) is it appropriate to calculate a standard deviation?