Ch1 - Descriptive Statistics

Tuesday, May 9, 2023 4:42 PM

Topics:

1. [2/2 Points] DETAILS PREVIOUS ANSWERS **DEVORESTAT9 1.1.006.** MY NOTES ASK YOUR TEACHER

The California State University (CSU) system consists of 23 campuses, from San Diego State in the south to Humboldt State near the Oregon border. A CSU administrator wishes to make an inference about the average distance between the hometowns of students and their campuses. Describe and discuss several different sampling methods that might be employed. (Select all that apply.)

- [x] The sample could be generated by taking a stratified random sample by taking a simple random sample from each of the 23 campuses and again asking each student in the sample to report the distance from their hometown to campus.
- [ ] Instead of taking a random sample, every student should be included in the study.
- [x] Certain problems arise with self reporting of distances, such as recording error or poor recall.
- [x] One could take a simple random sample of students from all students in the California State University system and ask each student in the sample to report the distance from their hometown to campus.
- [ ] There are no potential problems with self reporting of distances.

Would this be an enumerative or an analytic study? Explain your reasoning.

- ( ) This study is analytic because there exists an infinite population of objects from which to sample.
- (•) This study is enumerative because there exists a finite, identifiable population of objects from which to sample.

Need Help? Read It

Analytical study: infinite population of objects to sample
Enumerative study: finite, identifiable population of objects to sample

Sampling frame: a list of individuals or objects to be sampled
Simple random sample: data collection that entails selecting individuals or objects from a frame
Stratified random sample: separating the population units into nonoverlapping groups and taking samples from each one

We could either take a stratified random sample (option 1) or a simple random sample (option 4), we cannot include every student in a sample (option 2).

Certain problems do arise when reporting distance (option 3), excluding option 5.

Because the population is finite and identifiable, this is an enumerative study.

2. [2/2 Points] DETAILS PREVIOUS ANSWERS **DEVORESTAT9 1.1.007.** MY NOTES ASK YOUR TEACHER

A certain city divides naturally into ten district neighborhoods. How might a real estate appraiser select a sample of single-family homes that could be used as a basis for developing an equation to predict appraised value from characteristics such as age, size, number of bathrooms, distance to the nearest school, and so on?

- (•) An appropriate method is to generate a simple random sample of all single-family homes in the city or a stratified random sample by taking a simple random sample from each of the ten district neighborhoods.
- ( ) An appropriate method is to randomly select which district to survey, and record data for each home in that district.
- ( ) An appropriate method is to sample the most expensive homes in a randomly selected district.
- ( ) An appropriate method is to sample the newest homes in all ten districts.

Is the study enumerative or analytic?

- (•) enumerative
- ( ) analytic

Need Help? Read It

Option 4 is bad because it only includes new homes.

Option 3 is bad because it only includes the most expensive homes.

Option 2 is bad because it only includes one district (a single subpopulation).

Option 1 is best because we either take a simple random sample or stratified random sample with no constraints.

3. [2/2 Points] DETAILS PREVIOUS ANSWERS **DEVORESTAT9 1.1.008.MI.** MY NOTES ASK YOUR TEACHER

The amount of flow through a solenoid valve in an automobile's pollution-control system is an important characteristic. An experiment was carried out to study how flow rate depended on three factors: armature length, spring load, and bobbin depth. Four different levels (low, fair, moderate, and high) of each factor were chosen, and a single observation on flow was made for each combination of levels.

(a) The resulting data set consisted of how many observations?

64 observations

(b) Is this an enumerative or analytic study? Explain your reasoning.

- (•) This is an analytic study because the data would be collected on an existing process. There is no sampling frame.
- ( ) This is an enumerative study because there is a finite population of objects from which to sample.

Need Help? Read It Watch It Master It

(a)
Three factors, four different level possibilities for each

Armature Length	Spring Load	Bobbin Depth
Low	Low	Low
Low	Fair	Low
Low	Low	Fair

Fair	Low	Low
Fair	Fair	Low
.	.	.
.	.	.
.	.	.
High	High	High

4 levels for Armature Length x 4 levels for Spring Load x 4 levels for Bobbin Depth = 4^3 = 64

(b)
Because we are trying to improve the flow rate of the solenoid valve in this study instead of on an unchanging population, we have an analytical study with no sampling frame.

4. [12/12 Points] DETAILS PREVIOUS ANSWERS **DEVORESTAT9 1.2.010.**

Flexural strength is a measure of a material's ability to resist failure in bending. The accompanying data are on flexural strength of concrete (in MegaPascal, MPa, where 1 Pa (Pascal) = 1.45×10^{-4} psi):

5.9	7.2	7.3	6.2	8.1	6.8	7.0	7.6	6.8	6.5	7.0	6.3	7.9	9.0
8.2	8.7	7.8	9.7	7.4	7.7	9.7	7.8	7.7	11.6	11.5	11.8	10.8	

(a) Construct a stem-and-leaf display of the data. (Enter numbers from smallest to largest separated by spaces. Enter NONE for stems with no values.)

Stems	Leaves
5	9
6	2 3 5 8 8
7	0 0 2 3 4 6 7 7 8 8 9
8	1 2 7
9	0 7 7
10	8
11	5 6 8

What appears to be a representative strength value?
7 MPa

Do the observations appear to be highly concentrated about the representative value or rather spread out?
- ○ highly concentrated around the representative value
- ● spread out with a large range

(b) Does the display appear to be reasonably symmetric about a representative value, or would you describe its shape in some other way?
- ○ reasonably symmetric
- ● positive skewness
- ○ negative skewness
- ○ completely random distribution

(c) Do there appear to be any outlying strength values?
- ○ Yes
- ● No

(d) What proportion of strength observations in this sample exceed 10 MPa? (Round your answer to two decimal places.)
0.15

Need Help? Read It

(a)
Our leaf plot is simply organizing our data from smallest to largest and seeing where it clusters based upon the leading digit (essentially the following below except just the leading digit)

Stems	Leaves
5	5.9
6	6.2 6.3 6.5 6.8 6.8
7	7.0 7.0 7.2 7.3 7.4 7.6 7.7 7.7 7.8 7.8 7.9
8	8.2 8.7
9	9.0 9.7 9.7
10	10.8
11	11.5 11.6 11.8

Our "representative value" is 7 because it occurs the most frequently (protruding out the most)

But our data is not concentrated largely around the center, it's spread out (8 has much fewer than the nearby representative value)

(b)
Our leaf plot has positive skewness (AKA right skewness), consider inverting the leaf plot to see this:

Stems	5	6	7	8	9	10	11
Leaves	9	2 3 5 8 8	0 0 2 3 4 6 7 7 8 8 9	1 2 7	0 7 7	8	5 6 8

(c)
There are no values (leaves) completely disconnected completely such that their stems do not border other leaves (meaning we don't skip any spaces in between values).

(d)
4 strength values greater than 10 MPa: 10.8, 11.5, 11.6, 11.8
(4 strength values greater than 10 MPa) / (27 total strength values) = 4 / 27 = 0.15

5. [11/11 Points] DETAILS PREVIOUS ANSWERS DEVORESTAT9 1.2.011.MI.

The accompanying specific gravity values describe various wood types used in construction.

0.32	0.35	0.36	0.36	0.37	0.38	0.40	0.40	0.40
0.41	0.41	0.42	0.42	0.42	0.42	0.42	0.43	0.44
0.45	0.46	0.46	0.47	0.48	0.48	0.49	0.53	0.54
0.54	0.55	0.58	0.63	0.66	0.66	0.67	0.68	0.78

Construct a stem-and-leaf display using repeated stems. (Enter numbers from smallest to largest separated by spaces. Enter NONE for stems with no values.)

Stem = tenths	Leaf = hundredths
3L	2
3H	5 6 6 7 8
4L	0 0 0 1 1 2 2 2 2 2 3 4
4H	5 6 6 7 8 8 9
5L	3 4 4
5H	5 8
6L	3
6H	6 6 7 8
7L	NONE
7H	8

Comment on any interesting features of the display. (Select all that apply.)

- [x] There is a reasonably large amount of variation in the data.
- [ ] The data value 0.32 is a possible outlier.
- [ ] There is a small amount of variation in the data.
- [ ] The stem-and-leaf display shows that 0.55 is a good representative value for the data.
- [x] The data value 0.78 is a possible outlier.
- [x] The stem-and-leaf display shows that 0.45 is a good representative value for the data.
- [x] The display is not symmetric and appears to be positively skewed.
- [ ] The display is symmetric.

Need Help? Read It Watch It Master It

We're only concerned with the rightmost digit in the stem-and-leaf plot.

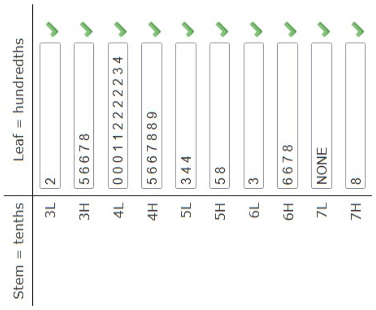
3L means 3-lower and 3H means 3-higher (0.32 is 3L, 0.35 is 3H, 0.38 is 3H, 4.0 is 4L, etc.)

Our data is very spread out so it has a large amount of variation.

0.32 is not a possible outlier because it's not disconnected or 'stranded' from other nearby leaves. However 0.78 is a possible outlier because it is not near any other leaves and is 'stranded'.

0.45 (around the 4L stem) is a good representative value because the most frequent values are here and the leaf plot bulges out the furthest here. This is not true for 0.55.

The display is positively skewed (right skewed) and therefore cannot be symmetric, this is visible by tilting the leaf plot:



6. [6/6 Points] DETAILS PREVIOUS ANSWERS DEVORESTAT9 1.2.019.S.

The number of contaminating particles on a silicon wafer prior to a certain rinsing process was determined for each wafer in a sample of size 100, resulting in the following frequencies:

Number of particles	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Frequency	1	3	3	12	11	15	19	10	10	4	5	3	1	2	1

USE SALT

(a) What proportion of the sampled wafers had at least one particle? At least five particles? (Round your answers to two decimal places.)

at least one particle 0.99

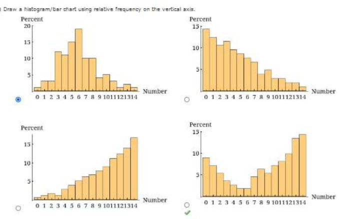
at least five particles 0.70

(b) What proportion of the sampled wafers had between five and ten particles, inclusive? Strictly between five and ten particles? (Round your answers to two decimal places.)

between five and ten particles, inclusive 0.63

strictly between five and ten particles 0.43

(c) Draw a histogram/bar chart using relative frequency on the vertical axis.



How would you describe the shape of the histogram?

- ( ) The histogram is negatively skewed and unimodal.
- ( ) The histogram is positively skewed and unimodal.
- (•) The histogram is fairly symmetric and unimodal.
- ( ) The histogram is fairly symmetric and bimodal.
- ( ) The histogram is positively skewed and bimodal.
- ( ) The histogram is negatively skewed and bimodal.

(a)

All but 1 out of 100 wafers has at least one particle (one wafer has 0 particles). Therefore 99% or 0.99 of wafers has at least one particle.

Number of wafers with at least five particles = (15 wafers w/five particles) + (19 wafers w/six particles) + (10 wafers w/seven particles) + (10 wafers w/eight particles) + (4 wafers w/nine particles) + (5 wafers w/ten particles) + (3 wafers w/eleven particles) + (1 wafer w/twelve particles) + (2 wafers w/thirteen particles) + (1 wafer/fourteen particles). There are 70 out of 30 wafers that have at least five particles. Therefore 70% or 0.70 of wafers have at least five particles.

15 + 19 + 10 + 10 + 4 + 5 + 3 + 1 + 2 + 1 = 70

(b)

Number of wafers with between five and ten particles (inclusive) = (15 wafers w/five particles) + (19 wafers w/six particles) + (10 wafers w/seven particles) + (10 wafers w/eight particles) + (4 wafers w/nine particles) + (5 wafers w/ten particles).
15 + 19 + 10 + 10 + 4 + 5 = 63

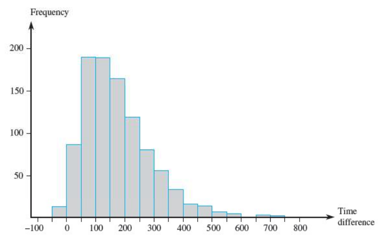
Number of wafers with between five and ten particles (restricted/exclusive) = (19 wafers w/six particles) + (10 wafers w/seven particles) + (10 wafers w/eight particles) + (4 wafers w/nine particles)
19 + 10 + 10 + 4 = 43

(c)

The frequency of the data is greatest at the center and only there as it spreads out further.
Therefore the data is symmetric and unimodal.

7. [3/3 Points] DETAILS PREVIOUS ANSWERS DEVORESTAT9 1.2.022. MY NOTES ASK YOUR TEACHER

How does the speed of a runner vary over the course of a marathon (a distance of 42.195 km)? Consider determining both the time to run the first 5 km and the time to run between the 35-km and 40-km points, and then subtracting the former time from the latter time. A positive value of this difference corresponds to a runner slowing down toward the end of the race. The accompanying histogram is based on times of runners who participated in several different Japanese marathons.†



What are some interesting features of this histogram? (Select all that apply.)

- [x] mostly positive values
- [x] positively skewed
- [ ] small sample size
- [ ] normally distributed
- [ ] small spread
- [ ] negatively skewed

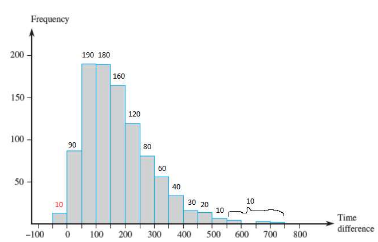
What is a typical difference value? (Enter an approximate value.)
100 sec

Roughly what proportion of the runners ran the late distance more quickly than the early distance? (Round your answer to two decimal places.)
0.01

The histogram is positively skewed (skewed rightwards) with mostly positive values.

The "typical" difference value is 100 sec because it's where most of the data is centered.

Only 1% or 0.01 of runners ran their last 35-km and 40-km faster than their first 5-km. This can be proven by the graph below.



Number of runners that ran their late distance more quickly than their early distance = 10 (in red)
Number of runners that ran their late distance slower than their early distance = 90 + 190 + 180 + 160 + 120 + 80 + 60 + 40 + 30 + 20 + 10 + 10 = 990

(Number of runners that ran their late distance more quickly than their early distance) / (number of runners that ran their late distance slower than their early distance) = 10 / 990 = 0.01

8. [15/15 Points] DETAILS PREVIOUS ANSWERS DEVORESTAT9 1.3.034.S.

Exposure to microbial products, especially endotoxin, may have an impact on vulnerability to allergic diseases. The following are data on concentration (EU/mg) in settled dust for one sample of urban homes and another of farm homes.

U: 6.0 5.0 11.0 33.0 4.0 5.0 80.0 18.0 35.0 17.0 23.0
F: 2.0 12.0 12.0 6.0 9.0 8.0 5.0 18.0 7.0 9.1 24.0 9.7 2.0 3.0 0.8

USE SALT

(a) Determine the sample mean for each sample. (Round your answers to four decimal places.)
urban homes 21.5455 EU/mg
farm homes 8.5067 EU/mg

How do they compare?
- ○ The average endotoxin concentration in farm homes is more than double the average concentration in urban homes.
- ○ The average endotoxin concentration is about the same in both urban and farm homes.
- ◉ The average endotoxin concentration in urban homes is more than double the average concentration in farm homes.

(b) Determine the sample median for each sample.
urban homes 17 EU/mg
farm homes 8 EU/mg

How do they compare?
- ○ The median endotoxin concentration is about the same in both urban and farm homes.
- ○ The median endotoxin concentration in farm homes is roughly double the median concentration in urban homes.
- ◉ The median endotoxin concentration in urban homes is roughly double the median concentration in farm homes.

Why is the median for the urban sample so different from the mean for that sample?
- ◉ The mean and median for urban homes are so different because the few large values raise the mean but not the median.
- ○ The mean and median for urban homes are so different because there are fewer observations.
- ○ The mean and median for urban homes are so different because the measure different aspects of the distribution.

(c) Calculate the trimmed mean for each sample by deleting the smallest and largest observation. (Round your answers to four decimal places.)
urban homes 17 EU/mg
farm homes 7.9077 EU/mg

What are the corresponding trimming percentages? (Round your answers to two decimal places.)
urban homes 9.09 %
farm homes 6.67 %

How do the values of these trimmed means compare to the corresponding means and medians?
Urban homes:
The trimmed mean is [less than] the mean of the entire sample. The trimmed mean is [equal to] the median of the entire sample.

Farm homes:
The trimmed mean is [less than] the mean of the entire sample. The trimmed mean is [less than] the median of the entire sample.

(a)
Sample size of urban homes = 11
Sample mean of urban homes = (sum of urban home concentration values) / (sample size of urban homes)
= (6.0 + 5.0 + 11.0 + 33.0 + 4.0 + 5.0 + 80.0 + 18.0 + 35.0 + 17.0 + 23.0) / 11 = 21.5455

Sample size of farm homes = 15
Sample mean of farm homes = (sum of farm home concentration values) / (sample size of farm homes)
= (2.0 + 12.0 + 12.0 + 6.0 + 9.0 + 8.0 + 5.0 + 18.0 + 7.0 + 9.1 + 24.0 + 9.7 + 2.0 + 3.0 + 0.8) / 15 = 8.5067

21.5455 (sample mean of urban homes) is more than double 8.5067 (sample mean of farm homes)

(b)
Median position of urban homes = (sample size of urban homes) / 2
= 11 / 2 = 5.5 = 5 (truncated)
Sorted concentrations of urban homes:
4 5 5 6 11 17 18 23 33 35 80
0 1 2 3 4 5 (positions)
Median of urban homes = 17

Median position of farm homes = (sample size of farm homes) / 2
= 15 / 2 = 7.5 = 7 (truncated)
Sorted concentrations of farm homes:
0.8 2 2 3 5 6 7 8 9 9.1 9.7 12 12 18 24
0 1 2 3 4 5 6 7 (positions)
Median of farm homes = 8

17 (sample median of urban homes) is more than double 8 (sample median of farm homes)

The sample mean for urban homes (21.5455) is much larger than the sample median for urban homes (17) because several large values raise the sample mean (which includes the sum of all values) but not the sample median (which is only the centermost value, unaffected by values).

(c)

Smallest value for urban homes: 4
Largest value for urban homes: 80
Urban homes sample size after trimming = 11 - 2 = 9
Trimmed urban homes mean after removing the largest and smallest values = (6.0 + 5.0 + 11.0 + 33.0 + 5.0 + 18.0 + 35.0 + 17.0 + 23.0) / 9 = 17

Smallest value for farm homes: 0.8
Largest value for farm homes: 24
Farm homes sample size after trimming = 15 - 2 = 13
Trimmed farm homes mean after removing the largest and smallest values = (2.0 + 12.0 + 12.0 + 6.0 + 9.0 + 8.0 + 5.0 + 18.0 + 7.0 + 9.1 + 9.7 + 2.0 + 3.0) / 13 = 7.9077

Urban homes trim percentage = (number of values removed from either largest or smallest end of data) / (original urban homes sample size) = 1 / 11 = 0.0909 = 9.09%

Farm homes trim percentage = (number of values removed from either largest or smallest end of data) / (original farm homes sample size) = 1 / 15 = 0.0667 = 6.67%

The urban homes trimmed mean of 17 is less than the urban homes sample mean of 21.5455 but is equal to the urban homes median of 17. The trimmed mean is less than the sample mean because we removed the larger value outlier.

The farm homes trimmed mean of 7.9077 is less than the farm homes sample mean of 8.5067 and less than the farm homes median of 8. This is because we removed a larger value that lowered the trimmed mean.

9. [1/1 Points] DETAILS PREVIOUS ANSWERS DEVORESTAT9 1.3.037. MY NOTES ASK YOUR TEACHER

The article "Snow Cover and Temperature Relationships in North America and Eurasia"† used statistical techniques to relate the amount of snow cover on each continent to average continental temperature. Data presented there included the following ten observations on October snow cover for Eurasia during the years 1970-1979 (in million km^2):

6.5 12.0 14.9 10.0 10.7 7.9 21.9 12.5 14.5 9.2

What would you report as a representative, or typical, value of October snow cover for this period, and what prompted your choice?
- ○ The median of this sample because the median is always the best central measure.
- ○ The mean of this sample because a potential outlier may produce a misleading median.
- ◉ The median of this sample because a potential outlier may produce a misleading mean.
- ○ The mean of this sample because the mean is always the best central measure.

Need Help? Read It

The median and mean are not always the best central measures, it depends on the circumstances (excluding options 1 and 4).

The median, not the mean, takes the centermost value and ignores potential outliers (whereas mean calculates as an average of all values, including outliers unless trimmed), excluding option 2, leaving option 3 as the only correct choice.

10. [3/3 Points] DETAILS PREVIOUS ANSWERS DEVORESTAT9 1.3.038.S. MY NOTES ASK YOUR TEACHER PRACTICE ANOTHER

Blood pressure values are often reported to the nearest 5 mmHg (100, 105, 110, etc.). The actual blood pressure values for nine randomly selected individuals are given below.

118.6 127.4 138.4 130.0 113.7 122.0 108.3 131.5 133.2

USE SALT

(a) What is the median of the *reported* blood pressure values?

125 ✓ mmHg

(b) Suppose the blood pressure of the second individual is 127.6 rather than 127.4 (a small change in a single value). What is the new median of the reported values?

130 ✓ mmHg

What does this say about the sensitivity of the median to rounding or grouping in the data?

- (●) When there is rounding or grouping, the median can be highly sensitive to small change.
- ( ) When there is rounding or grouping, the median is only sensitive to large changes.
- ( ) When there is rounding or grouping, the median is not sensitive to small changes.

✓

Need Help? Read It

Sample size = 9
Median position = (sample size) / 2 = 9 / 2 = 4.5 = 4 (truncated)

Reported (grouped) values of blood pressures (rounded to nearest 5mmHg) and sorted:
110 115 120 120 125 130 130 135 140
0 1 2 3 4 (position)

The median of the reported blood pressure values = 125

If we round 127.6 down to 127.4, the new reported (grouped) values of blood pressures, sorted, are:
110 115 120 120 130 130 130 135 140
0 1 2 3 4 (position)

The new median reported blood pressure value = 130

This proves that when rounding or grouping data, the median can be highly sensitive to small changes.

11. [11/11 Points] DETAILS PREVIOUS ANSWERS DEVORESTAT9 1.4.045.MI. MY NOTES ASK YOUR TEACHER PRACTICE ANOTHER

Suppose the value of Young's modulus (GPa) was determined for cast plates consisting of certain intermetallic substrates, resulting in the following sample observations.

116.8	115.7	114.5	115.2	115.6

(a) Calculate $\bar{x}$ (in GPa).

115.56 ✓ GPa

Calculate the deviations from the mean.

x	116.8	115.7	114.5	115.2	115.6
deviation	1.24 ✓	0.14 ✓	-1.06 ✓	-0.36 ✓	0.04 ✓

(b) Use the deviations calculated in part (a) to obtain the sample variance (in GPa^2).

s^2 = 0.703 ✓ GPa^2

Use the deviations calculated in part (a) to obtain the sample standard deviation (in GPa). (Round your answer to three decimal places.)

s = 0.8385 ✓ GPa

(c) Calculate s^2 (in GPa^2) by using the computational formula for the numerator S_{xx}.

0.703 ✓ GPa^2

(d) Subtract 100 from each observation to obtain a sample of transformed values. Now calculate the sample variance (in GPa^2) of these transformed values.

0.703 ✓ GPa^2

Compare it to s^2 for the original data.

- ( ) The variance in part (d) is greater than the variance in part (b).
- (●) The variance in part (d) is equal to the variance in part (b).
- ( ) The variance in part (d) is smaller than the variance in part (b).

✓

(a)

Sample mean = x̄ = (116.8 + 115.7 + 114.5 + 115.2 + 115.6) / 5 = 115.56

Sample deviation = (x - x̄)

x	116.8	115.7	114.5	115.2	115.6
(x - x̄)	116.8 - 115.56 = 1.24	115.7 - 115.56 = 0.14	114.5 - 115.56 = -1.06	115.2 - 115.56 = -0.36	115.6 - 115.56 = 0.04

(b)

Sample variance = (x - x̄)^2

x	116.8	115.7	114.5	115.2	115.6
(x - x̄)	116.8 - 115.56 = 1.24	115.7 - 115.56 = 0.14	114.5 - 115.56 = -1.06	115.2 - 115.56 = -0.36	115.6 - 115.56 = 0.04
(x - x̄)^2	(1.24)^2 = 1.5376	(0.14)^2 = 0.0196	(-1.06)^2 = 1.1236	(-0.36)^2 = 0.1296	(0.04)^2 = 0.0016

Sample variance = s^2 = (sum of the sample variances) / (sample size - 1) = ∑[x - x̄]^2 / (n - 1) = (1.5376 + 0.0196 + 1.1236 + 0.1296 + 0.0016) / (5-1) = 0.703

Sample standard deviation = s = Sqrt(s^2) = Sqrt(sample variance) = Sqrt(0.703) = 0.8385

(c)

Alternative method to calculate sample variance and sample standard deviation using the sum of squared deviations from the sample mean (or S_{xx})

x	116.8	115.7	114.5	115.2	115.6
(x^2)	(116.8)^2 = 13,642.24	(115.7)^2 = 13,386.49	(114.5)^2 = 13,110.25	(115.2)^2 = 13,271.04	(115.6)^2 = 13,363.36

∑[x] = 116.8 + 115.7 + 114.5 + 115.2 + 115.6 = 577.8
∑[x]^2 = 13,642.24 + 13,386.49 + 13,110.25 + 13,271.04 + 13,363.36 = 66,773.38

S_{xx} = sum of squared deviations from the sample mean = ∑[x - x̄]^2 = ∑[x]^2 - (∑[x])^2 / n
= (66,773.38) - (577.8)^2 / 5 = 2.812

Sample variance = s^2 = (sum of squared deviations from the sample mean) / (sample size - 1) = (S_{xx}) / (n - 1)
= (2.812) / (5 - 1) = 0.703

(d)

x	116.8	115.7	114.5	115.2	115.6
(x - 100)	116.8 - 100 = 16.8	115.7 - 100 = 15.7	114.5 - 100 = 14.5	115.2 - 100 = 15.2	115.6 - 100 = 15.6
(x - 100)^2	(16.8)^2 = 282.24	(15.7)^2 = 246.49	(14.5)^2 = 210.25	(15.2)^2 = 231.04	(15.6)^2 = 243.36

∑[x - 100] = 16.8 + 15.7 + 14.5 + 15.2 + 15.6 = 77.8
∑[x - 100]^2 = 282.24 + 246.49 + 210.25 + 231.04 + 243.36 = 1,213.38

S_{xx} = sum of squared deviations from the sample mean = ∑[x - x̄]^2 = ∑[x]^2 - (∑[x])^2 / n
= (1,213.38) - (77.8)^2 / 5 = 2.812

Sample variance = s^2 = (sum of squared deviations from the sample mean) / (sample size - 1) = (S_{xx}) / (n - 1)
= (2.812) / (5 - 1) = 0.703

Changing each observation consistently (<u>when adding or subtracting</u>) will yield the same variance and thus the same deviations for the transformed values.

12. [4/4 Points] DETAILS PREVIOUS ANSWERS DEVORESTAT9 1.4.051.S. MY NOTES ASK YOUR TEACHER PRACTICE ANOTHER

An article reported the following data on oxidation-induction time (min) for various commercial oils:

86 104 130 160 180 195 135 145 213 105 145
151 154 136 87 99 92 119 129

USE SALT

(a) Calculate the sample variance and standard deviation. (Round your answers to four decimal places.)

s^2 = 1,287.7778 ✔ min^2
s = 35.8856 ✔ min

(b) If the observations were reexpressed in hours, what would be the resulting values of the sample variance and sample standard deviation? Answer without actually performing the reexpression. (Round your answer to four decimal places.)

s^2 = 0.3577 ✔ hr^2
s = 0.5981 ✔ hr

Need Help? Read It Watch It

(a)

1	oxidation-induction time (min)	Squared Values
2	86	7396
3	104	10816
4	130	16900
5	160	25600
6	180	32400
7	195	38025
8	135	18225
9	145	21025
10	213	45369
11	105	11025
12	145	21025
13	151	22801
14	154	23716
15	136	18496
16	87	7569
17	99	9801
18	92	8464
19	119	14161
20	129	16641
21		
22	**Sum**	2565
23	**Sum of Squares**	369455

Sample size = 19

S_{xx} = sum of squared deviations from the sample mean = ∑[x - x̄]^2 = ∑[x]^2 - (∑[x])^2 / n
= (369,455) - (2,565)^2 / 19 = 23,180

Sample variance = s^2 = (sum of squared deviations from the sample mean) / (sample size - 1) = (S_{xx}) / (n - 1)
= 23,180 / (19 - 1) = 1,287.7778

Sample standard deviation = s = Sqrt(s^2) = Sqrt(sample variance) = Sqrt(1,287.7778) = 35.8856

(b)

	A	B	C
1	oxidation-induction time (min)	oxidation-induction time (hr)	Squared Values
2	86	1.433333333	2.054444444
3	104	1.733333333	3.004444444
4	130	2.166666667	4.694444444
5	160	2.666666667	7.111111111
6	180	3	9
7	195	3.25	10.5625
8	135	2.25	5.0625
9	145	2.416666667	5.840277778
10	213	3.55	12.6025
11	105	1.75	3.0625
12	145	2.416666667	5.840277778
13	151	2.516666667	6.333611111
14	154	2.566666667	6.587777778
15	136	2.266666667	5.137777778
16	87	1.45	2.1025
17	99	1.65	2.7225
18	92	1.533333333	2.351111111
19	119	1.983333333	3.933611111
20	129	2.15	4.6225
21			
22	**Sum**	42.75	
23	**Sum of Squares**	102.6263889	
24		6.438888889	

S_{xx} = sum of squared deviations from the sample mean = ∑[x - x̄]^2 = ∑[x]^2 - (∑[x])^2 / n
(102.6264) - (42.75)^2 / 19 = 6.4389

Sample variance = s^2 = (sum of squared deviations from the sample mean) / (sample size - 1) = (S_{xx}) / (n - 1)
= 6.4389 / (19 - 1) = 0.3577

Sample standard deviation = s = Sqrt(s^2) = Sqrt(sample variance) = Sqrt(0.3577) = 0.5981

Unlike the previous problem, changing each observation consistently <u>when multiplying or dividing</u> will yield different variances and deviations from the transformed values.