Solution Section 1.9 – Measures of Dispersion

Exercise

In statistics, how do variation and variance differ?

Solution

Variation is a general descriptive term that refers to the fact that all the items being measured are not identical. In Statistics, variance is a specific and well-defined measure of variation the has a particular mathematical formula.

Exercise

Collegiate Dictionary has 1459 pages of defined words. Listed below are the numbers of defined words per page for a simple random sample of those pages. If we use this sample as a basis for estimating the total number of defined words in the dictionary.

51 63 36 43 34 62 73 39 53 79

a) Find the range, variance, and standard deviation.

b) How does the variation of these numbers affect our confidence on the accuracy of the estimate?

Solution

a) $\bar{x} = \frac{\sum x}{n} = \frac{533}{10} = 53.3$

$Range = 79 - 34 = 45$ **words**

$Variance: = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{2206.10}{9}$

$= 245.12 \ words$

$Standard\ deviation: \sqrt{245.12} = 15.7 \ words$

x	$x-\bar{x}$	$(x-\bar{x})^2$	x^2
34	-19.3	372.49	1156
36	-17.3	299.29	1296
39	-14.3	204.49	1521
43	-10.3	106.09	1849
51	-2.3	5.29	2601
53	-0.3	0.09	2809
62	8.7	75.69	3844
63	9.7	94.09	3969
73	19.7	388.09	5329
79	25.7	660.49	6241
533	0.0	2206.10	30615

b) There seems to be considerable variation from page to page. For small samples with n = 10, there could be considerable variation in the sample mean and, therefore, considerable variation in the projected totals for the entire dictionary. It appears that there would be a question about the accuracy of an estimate based on this sample for the total number of words.

Exercise

The National Highway Traffic Administration conducted crash tests of child booster seats for cars. Listed below are results from those teats, with the measurements given in hic (standard head injury condition units.

774 649 1210 546 431 612

a) Find the range, variance, and standard deviation

b) According to the safety requirement, the hic measurement should be less than 1000 hic. Do the results suggest that all of the child booster seats meet the specified requirement?

Solution

x	x^2
431	185761
546	298116
612	374544
649	421201
774	599076
1210	1464100
4222	3342798

a) $\bar{x}=\frac{\sum x}{n}=\frac{774+649+1210+546+431+612}{6}\ \boxed{=703.67}$

$Range = 1210 - 431 = 779$ ***hic***

$$Variance: \frac{n\left(\sum x^2\right)-\left(\sum x\right)^2}{n(n-1)}=\frac{6(3342798)-(4222)^2}{6(5)}$$

$$\boxed{=74383.5\ \ hic^2}$$

$Standard\ deviation: \sqrt{74383.5}\ \boxed{=272.7\ \ hic}$

b) Yes, there seems to be much variation – mainly due to the largest value, which appears to be substantially different from the others.

Exercise

The insurance Institution for Highway Safety conducted tests with crashes of new cars traveling at 6 mi/h. The total cost of the damages was found for a simple random sample of the tested cars and listed below

\$7448 \$4911 \$9051 \$6374 \$4277

a) Find the range, variance, and standard deviation

b) Do the different measures of center differ very much?

Solution

x	x^2
4277	18292729
4911	24117921
6374	40627876
7448	55472704
9051	81920601
32061	220431831

a) $Range = 9051 - 4277 = \$4744$

$$Variance: \frac{n\left(\sum x^2\right)-\left(\sum x\right)^2}{n(n-1)}=\frac{5(220431831)-(32061)^2}{5(4)}$$

$$\boxed{=3712571.7\ \ dollars^2}$$

$Standard\ deviation: \sqrt{3712571.7}\ \boxed{=\$1926.8}$

b) A value is considered unusual if it differs from the mean by more than two standard deviations. Since \$10,000 differs from the mean by $\frac{10,000-6,412.2}{1,926.8}=1.86$ standard deviations, in this context it would not be considered an unusual value.

Exercise

Listed below are the durations (in hours) of a simple random sample of all flights (as of this writing) of NASA's Space Transport System (space shuttle).

73 95 235 192 165 262 191 376 259 235 381 331 221 244 0

a) Find the range, variance, and standard deviation

b) How might that duration time be explained?

Solution

a) *Range* = 381 – 0 = 381 ***hrs***.

Variance: $\frac{n\left(\sum x^2\right)-\left(\sum x\right)^2}{n(n-1)}=\frac{15(865574)-(3260)^2}{15(14)}$

$=11219.1\ hrs^2$

Standard deviation: $\sqrt{11219.1}=105.9\ hrs$

$\bar{x}=\frac{\sum x}{n}=\frac{3260}{15}=217.3$

x	x^2
0	0
73	5329
95	9025
165	27225
191	36481
192	36864
221	48841
235	55225
235	55225
244	59536
256	65536
262	68644
331	109561
376	141376
381	145161
3260	865574

b) A value is considered unusual if it differs from the mean by more than two standard deviations. Since 0 hours differs from the mean by $\frac{0-217.3}{105.9}=-2.05$ standard deviations, in this context it would not be considered an unusual value.

Exercise

Listed below are the playing times (in seconds) of songs that were popular at the time of this writing.

448 242 231 246 246 293 280 227 213
262 239 213 258 255 257 244

a) Find the range, variance, and standard deviation

b) Is there on time that is very different from the others?

x	x^2
213	45369
213	45369
227	51529
231	53361
239	57121
242	58564
244	59536
246	60516
246	60516
255	65025
257	66049
258	66564
262	68644
280	78400
293	85849
448	200704
4154	1123116

Solution

a) *Range* $= 448 - 213 = 235$ ***sec***

$$\text{Variance: } s^2 = \frac{n\left(\sum x^2\right) - \left(\sum x\right)^2}{n(n-1)}$$

$$= \frac{16(1123116) - (4154)^2}{16(15)} \qquad \left(16*1123116-4154^2\right) / (16*15)$$

$$= 2975.6 \ \text{sec}^2$$

Standard deviation: $\sqrt{2975.6} = 54.5 \ \textit{sec}$

$$\overline{x} = \frac{\sum x}{n} = \frac{4154}{16} = 259.6$$

b) If the highest playing time is omitted, then

$$\overline{x} = \frac{3706}{15} = 247.1 \qquad s = \frac{n\left(\sum x^2\right) - \left(\sum x\right)^2}{n(n-1)} = \sqrt{\frac{15(922412) - (3706)^2}{15(14)}} = 22.0 \ \text{sec}.$$

The change is substantial.

Exercise

Listed below are numbers of satellites in orbit from different countries.

158 17 15 17 7 3 5 1 8 3 4 2 4 1 2 3 1 1 1 1 1 1 1 1

a) Find the range, variance, and standard deviation

b) Does on country have an exceptional number of satellites?

Solution

a) *Range* = 158 – 1 = 157 ***satellites***

$$Variance:\ s^2 = \frac{n\left(\sum x^2\right) - \left(\sum x\right)^2}{n(n-1)}$$

$$= \frac{24(26017) - (259)^2}{24(23)}$$

$\left(24*26017-259^2\right) / (24*23)$

$$= 1009.7\ satellites^2$$

Standard deviation: $s = \sqrt{1009.7} = 31.8\ satellites$

$$\overline{x} = \frac{\sum x}{n} = \frac{259}{24} = 10.8$$

b) A values is considered unusual if it differs from the mean by more than two standard deviations. Since 0 satellites differs from the mean by $\frac{0-10.8}{31.8} \approx -0.34$ standard deviations, in this context it would not be considered an unusual value.

There are many countries with 0 satellites, and since the original data included no 0's, it appears that the data was gathered only from countries that had at least one satellite.

x	x^2
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
2	4
2	4
3	9
3	9
3	9
4	16
4	16
5	25
7	49
8	64
15	225
17	289
18	328
158	24964
259	26017

Exercise

Listed below are costs (in dollars) of roundtrip flights from JFK airport in NY City to San Francisco. (All flights involve one stop and a two-week stay.) The airlines are US Air, Continental, Delta, United, American, Alaska, and Northwest.

30 Days in Advance	244	260	264	264	278	318	280
1 Day in Advance	456	614	567	943	628	1088	536

Find the coefficient of variation for each of the two sets of data, then compare the variation.

Solution

x	x^2	x	x^2
244	59536	456	207936
260	67600	536	287296
264	69696	567	321489
264	69696	614	376996
278	77284	628	394384
280	78400	943	889249
318	101124	1088	1183744
1908	523336	4832	3661094

For 30 days:

$$\bar{x} = \frac{\sum x}{n} = \frac{1908}{7} \underline{= \$272.6}$$

$$s^2 = \frac{n\left(\sum x^2\right) - \left(\sum x\right)^2}{n(n-1)} = \frac{7(523336) - (1908)^2}{7(6)} \underline{= 545.0 \ \ dollars^2}$$

$$s = \sqrt{545} \underline{= \$23.3}$$

Coefficient of variation: $CV = \frac{s}{\bar{x}} = \frac{23.3}{272.6} \approx 0.086 \underline{= 8.6\%}$

For 1 day:

$$\bar{x} = \frac{\sum x}{n} = \frac{4832}{7} \underline{= \$690.3}$$

$$s^2 = \frac{n\left(\sum x^2\right) - \left(\sum x\right)^2}{n(n-1)} = \frac{7(3661094) - (4832)^2}{7(6)} \underline{= 54272.2 \ \ dollars^2}$$

$$s = \sqrt{54272.2} \underline{= \$233.0}$$

Coefficient of variation: $CV = \frac{s}{\bar{x}} = \frac{233.0}{690.3} \approx 0.337 \underline{= 33.7\%}$

There is considerably more variation among the costs for tickets purchased one day in advance.

Exercise

The trend of thinner Miss America winners has generated charges that the contest encourages unhealthy diet habits among young women. Listed below are body mass indexes (BMI) for Miss America winners from two different periods.

BMI (1920 – 1930)	20.4	21.9	22.1	22.3	20.3	18.8	18.9	19.4	18.4	19.1
BMI – (*from recent winners*)	19.5	20.3	19.6	20.2	17.8	17.9	19.1	18.8	17.6	16.8

Find the coefficient of variation for each of the two sets of data, then compare the variation.

Solution

x	x^2		x	x^2
18.4	338.56		16.8	282.24
18.8	353.44		17.6	309.76
18.9	357.21		17.8	316.84
19.1	364.81		17.9	320.41
19.4	376.36		18.8	353.44
20.3	412.09		19.1	364.81
20.4	416.16		19.5	380.25
21.9	479.61		19.6	384.16
22.1	488.41		20.2	408.04
22.3	497.29		20.3	412.09
201.6	4083.94		187.6	3532.04

For BMI (1920 – 1930):

$$\bar{x} = \frac{\sum x}{n} = \frac{201.6}{10} \underline{= 20.16}$$

$$s^2 = \frac{n\left(\sum x^2\right) - \left(\sum x\right)^2}{n(n-1)} = \frac{10(4083.94) - (201.6)^2}{10(9)} \underline{= 2.19}$$

$$s = \sqrt{2.19} \underline{= 1.48}$$

Coefficient of variation: $CV = \frac{s}{\bar{x}} = \frac{1.48}{20.16} \approx 0.073 \underline{= 7.3\%}$

For BMI – (from recent winners):

$$\bar{x} = \frac{\sum x}{n} = \frac{187.6}{10} \underline{= 18.76}$$

$$s^2 = \frac{n\left(\sum x^2\right) - \left(\sum x\right)^2}{n(n-1)} = \frac{10(3532.04) - (187.6)^2}{10(9)} \underline{= 1.41}$$

$$s = \sqrt{1.41} \underline{= 1.19}$$

Coefficient of variation: $CV = \frac{s}{\bar{x}} = \frac{1.19}{18.76} \approx 0.063 \underline{= 6.3\%}$

The BMI values for the two time period appear to exhibit about the same amount of variation.

Exercise

Find the Standard Deviation from the frequency distribution and find the standard deviation of sample summarized in a frequency distribution table by using the formula $s=\sqrt{\dfrac{n\left[\sum\left(f\cdot x^2\right)\right]-\left[\sum(f\cdot x)\right]^2}{n(n-1)}}$, where x represents the class midpoint, f represents the class frequency, and n represents the total number of sample values.

a)

Tar (mg) in Nonfiltered Cigarettes	***Frequency***
10 – 13	1
14 – 17	0
18 – 21	15
22 – 25	7
26 – 29	2

b)

Pulse Rates of Females	***Frequency***
60 – 69	12
70 – 79	14
80 – 89	11
90 – 99	1
100 – 109	1
110 – 119	0
120 – 129	1

Solution

a) The x values are the class midpoints from the given frequency table.

x	f	$f\cdot x$	$f\cdot x^2$
11.5	1	11.5	132.25
15.5	0	0	0
19.5	15	292.5	5703.75
23.5	7	164.5	3867.75
27.5	2	55.0	1512.50
	25	523.5	11214.25

$$s^2=\frac{n\left(\sum x^2\right)-\left(\sum x\right)^2}{n(n-1)}=\frac{25(11214.25)-(523.5)^2}{25(24)}=10.507$$

$$s=\sqrt{10.507}=3.2\ mg$$

This is the same as the true value of 3.2 mg.

b) The x values are the class midpoints from the given frequency table.

x	f	$f\cdot x$	$f\cdot x^2$
64.5	12	774.0	49923.00
74.5	14	1043.0	77703.5
84.5	11	929.5	78542.75
94.5	1	94.5	8930.25
104.5	1	104.5	10920.25
114.5	0	0	0
124.5	1	124.5	15500.25
	40	3070.0	241520.00

$$s^2 = \frac{n\left(\sum x^2\right) - \left(\sum x\right)^2}{n(n-1)} = \frac{40(241520) - (3070)^2}{40(39)} \underline{=151.218}$$

$$s = \sqrt{151.218} \ \underline{= 12.3 \quad beats/min}$$

This is close to the true value of 12.5 beats/min.

Exercise

Heights of women have a bell-shaped distribution with a mean of 161 cm and a standard deviation of 7 cm. Using the empirical rule, what is the approximate percentage of women between

a) 154 cm and 168 cm?
b) 147 cm and 175 cm?

Solution

a) The range from 154 to 168 is $\overline{x} \pm 1s$ $(161-7=154 \quad 161+7=168)$. The empirical rule suggests that about 68% of the data values should fall within those limits.

b) The range from 147 to 175 is $\overline{x} \pm 2s$ $(161-14=147 \quad 161+14=175)$. The empirical rule suggests that about 95% of the data values should fall within those limits.

Exercise

The author's Generac generator produces voltage amounts with a mean of 125.0 volts and standard deviation of 0.3 volts, and the voltages have a bell-shaped distribution. Using the empirical rule, what is the approximate percentage of voltage amounts between

a) 124.4 volts and 125.6 volts?
b) 124.1 volts and 125.9 volts?

Solution

a) The range from 124.4 to 125.6 is $\overline{x} \pm 2s$ $(125.0-2(.3)=124.4 \quad 125.0+2(.3)=125.6)$. The empirical rule suggests that about 95% of the data values should fall within those limits.

b) The range from 124.1 to 125.9 is $\overline{x} \pm 3s$ $(125.0-3(.3)=124.3 \quad 125.0+3(.3)=125.9)$. The empirical rule suggests that about 99.7% of the data values should fall within those limits.

Exercise

The mean value of land and buildings per acre from a sample of farms is $1,500, with a standard deviation of $200. Using the empirical rule, estimate the percent of farms whose land and building values per acre are between $1,300 and $1,700. (Assume the data set has a bell-shaped distribution.)

Solution

$$(1300,\ 1700) \to (1500 - 1(200),\ \ 1500 + 1(200))$$

$$\Rightarrow\ 1500 \pm 200 = \overline{x} \pm \delta$$

68% of the farms have values between $1,300 and $1,700 per acre

Exercise

The mean value of land and buildings per acre from a sample of farms is $2,400, with a standard deviation of $450. Using the empirical rule, between what two values do about 95% of the data lie? (Assume the data set has a bell-shaped distribution.)

Solution

$$95\% \quad \to \quad x \pm 2\delta$$

$$2400 \pm 2(450) \quad \to \begin{cases} 2400 - 900 = 1500 \\ 2400 + 900 = 3300 \end{cases}$$

95% of the farms have values between $1,500 and $3,300 per acre

Exercise

Heights of women have a bell-shaped distribution with a mean of 161 cm and a standard deviation of 7 cm. Using Chebyshev's Theorem, what do we know about the percentage of women with heights that are within 2 standard deviations of the mean? What are the minimum and maximum heights that are within 2 standard deviations of the mean?

Solution

Chebyshev's Theorem states that for any set of data there must be at least $1 - \left(\frac{1}{2}\right)^2 = 1 - \frac{1}{4} = \frac{3}{4}$ (75%) of the data values within 2 standard deviations of the mean. In this context, the limits are $161 \pm 2(7)$: there must be at least 75% of the heights between 147 cm and 175 cm.

Exercise

The author's Generac generator produces voltage amounts with a mean of 125.0 volts and standard deviation of 0.3 volts. Using Chebyshev's theorem, what do we know about the percentage of voltage amounts that are within 3 standard deviations of the mean? What are the minimum and maximum voltage amounts that are within 3 standard deviations of the mean?

Solution

Chebyshev's Theorem states that for any set of data there must be at least

$$1-\frac{1}{k^2}=1-\frac{1}{3^2}=1-\frac{1}{9}\ \underline{=\frac{8}{9}}|$$

(89%) of the data values within 3 standard deviations of the mean.

In this context, the limits are $125\pm 3(0.3)$: there must be at least 75% of the heights between 124.1 volts and 125.9 volts.

Exercise

The mean time in a women's 400-meter dash is 57.07 seconds, with a standard deviation of 1.05 seconds. Apply Chebyshev's Theorem to the data using $k=2$. Interpret the results.

Solution

$$1-\frac{1}{k^2}=1-\frac{1}{2^2}=1-\frac{1}{4}\ \underline{=0.75}|$$

At least 75% of the 400-meter dash within 2 standard deviations of the mean

The limits are $x\pm 2\delta \rightarrow 57.07\pm 2(1.05)\Rightarrow \begin{cases}57.07-2.1=54.97\\ 57.07+2.1=59.17\end{cases}$

There must be at least 75% of the 400-meter dash between 54.97 and 59.17 seconds.

Exercise

The number of gallons of water consumed per day by a small village are listed. Make a frequency distribution (using five classes) for the data set. Then approximate the population mean and the population standard deviation of the data set.

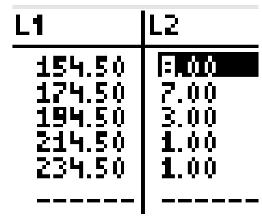
167	180	192	173	145	151	174	175	178	160
195	224	244	146	162	146	177	163	149	188

Solution

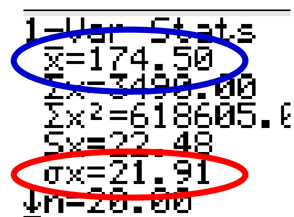
$$\text{Class width} = \frac{Max - Min}{5}$$

$$= \frac{244 - 145}{5}$$

$$= 19.8 \approx 20$$



Class	***x***	***f***	$f \cdot x$
145 – 164	154.5	8	1236.0
165 – 184	174.5	7	1221.5
185 – 204	194.5	3	583.5
205 – 224	214.5	1	214.5
225 – 244	234.5	1	234.5
		20	3490.0



$$\mu = \frac{\sum xf}{N} = \frac{3490}{20} = 174.5$$

$x-\mu$	$(x-\mu)^2$	f	$(x-\mu)^2 f$
–20	400	8	3200
0	0	7	0
20	400	3	1200
40	1600	1	1600
60	3600	1	3600
			9600

$$\delta = \sqrt{\frac{\sum (x-\mu)^2 f}{N}}$$

$$= \sqrt{\frac{9600}{20}}$$

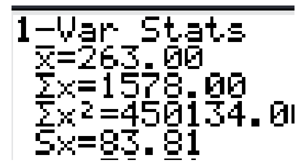
$$\approx 21.9$$

Exercise

To get the best deal on a microwave oven, Jeremy called six appliance stores and asked the cost of a specific model. The prices he was quoted are listed below:

$325 $384 $156 $210 $219 $284

Find the variance for the given data.



Solution

$\delta \approx 83.81$

$variance = 83.81^2 = \$7,024.00$

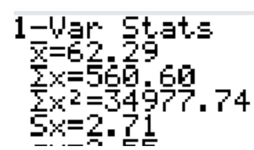
Exercise

Compare the variation in heights to the variation in weights of thirteen-year old girls. The heights (in inches) and weights (in pounds) of nine randomly selected thirteen-year old girls as listed below

Heights (inches):	59.3	61.2	62.6	64.7	60.1	58.3	64.6	63.7	66.1
Weights (pounds):	87	96	91	119	96	90	123	98	139

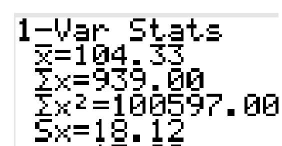
Find the coefficient of variation for each of the two sets of data, then compare the variation

Solution



Heights: $\bar{x} = 62.29, \quad s = 2.71$

Coefficient of variation: $CV = \frac{s}{\bar{x}} = \frac{2.71}{62.29} \approx 0.043 = 4.3\%$



Weights: $\bar{x} = 104.33, \quad s = 18.12$

Coefficient of variation: $CV = \frac{s}{\bar{x}} = \frac{18.12}{104.33} \approx 0.1736 = 17.4\%$

There is substantially more variation in the weights than in the heights of the girls

Exercise

The amount of Jen's monthly phone bill is normally distributed with a mean of $56 and a standard deviation of $9. What percentage of her phone bills are between $29 and $83? Use the empirical rule to solve.

Solution

$83 = 56 + 9 + 9 + 9$

$29 = 56 - 9 - 9 - 9$

29 and 83 are each exactly 3 standard deviation away from the mean 56.

The empirical rule tells us that about 99.7% of all values are within 3 standard deviation of the mean, so about 99.7% of her phone bills are between 70 and 130.