The "Standard Deviation" & the "Variance" ~ Definitional Formulas Review & FAQs

by
Anthony A. Walsh, Ph.D.
Version 2.2

25-February-2004

Notes on Statistical Notation

- Statistical symbol notation is not 100% standardized in the multitude of texts, software, & technical writing devoted to this subject.
- Therefore, the notation contained here will depict what is to be considered standard for this course.
- Furthermore, this notation scheme is intended to be consistent with that which is used in the textbook you are reading this semester.

The Definitional Formulas Used for Variation(SS),Variance(MS), & Standard Deviation in this class are Shown Below

$$SS_{samp} = \sum(X-\bar{X})^2 \qquad SS_{pop} = \sum(x-\mu)^2$$

A $$s^2 = \frac{\sum(X-\bar{X})^2}{n}$$

B $$s = \sqrt{\frac{\sum(X-\bar{X})^2}{n}}$$

C $$\mathscr{s}^2 = \frac{\sum(X-\bar{X})^2}{n-1}$$

D $$\mathscr{s} = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$$

E $$\sigma^2 = \frac{\sum(x-\mu)^2}{N}$$

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

Variation, SS, the "Sum of the Squares"

Variation, defined as the SUM OF THE SQUARES OF THE DEVIATION SCORES(or other values) FROM THE MEAN (**SS**), occurs only in descriptive form, i.e, it does not have an inferential counterpart. It will always just be the statistic defined above and nothing more. It does, however, have both a SAMPLE NOTATION VERSION(left below) and POPULATION NOTATION VERSION(right below) distinguished from each other only by the symbol used for the mean.

$$SS_{samp} = \sum(X-\bar{X})^2 \qquad SS_{pop} = \sum(x-\mu)^2$$

The Definitional Source for our Variation, Variances & Standard Deviations are Shown Here

	Original Data X	Deviation Values X-Mean		Deviation Values Squared (X-Mean)^2	
	58	10.778		116.160	
	63	15.778		248.938	
	22	-25.222		636.160	
	35	-12.222		149.383	
	50	2.778		7.716	
	43	-4.222		17.827	
	43	-4.222		17.827	
	53	5.778		33.383	
	58	10.778		116.160	
SUMs=	425			1343.556	=Variation, SS
Mean=	47.222	Variance	=	149.284	s^2 or σ^2
		StdDev	=	12.218	s or σ
N=	9				
		Variance	=	167.944	$\mathscr{s}^2$
		StdDev	=	12.959	$\mathscr{s}$
	Descriptive Statistics				Inferential Statistics

The "Descriptive Statistics" formulas use N or n in their calculation. The "Inferential Statistics" formulas use n-1.

Symbol Usage Summary

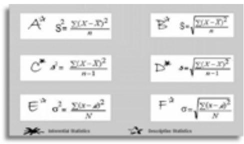
Some Symbols for Sample Statistics, Population Parameters

	n Sample Statistic	N Population Parameter
* *Mean*	$\bar{X}$	μ
* *Variance*	s^2	σ^2
* *Standard deviation*	s	σ
* **Estimated Pop Variance**	$\mathscr{s}^2$	
* **Estimated Pop Stand.Dev.**	$\mathscr{s}$	

* Descriptive Statistics * **Inferential Statistics (uses sample data)**

Question 1

- Which of the formulas to the right is a descriptive statistic calculating a standard deviation on the actual population data?



7

Question 1 Answer

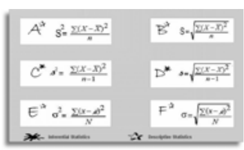
- Formula “F” to the right is a descriptive statistic calculating a standard deviation on the actual population data?

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

8

Question 2

- Which of the formulas to the right is a descriptive statistic calculating a variance on sample data?



9

Question 2 Answer

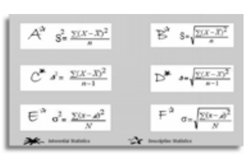
- Formula “A” to the right is a descriptive statistic calculating a variance on sample data?

A $S^2 = \frac{\sum(X-\bar{X})^2}{n}$

10

Question 3

- Which of the formulas to the right is an inferential statistic calculating an estimate of the population variance based on sample data?



11

Question 3 Answer

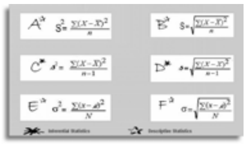
- Formula “C” to the right is an inferential statistic calculating an estimate of the population variance based on sample data?

C $s^2 = \frac{\sum(X-\bar{X})^2}{n-1}$

12

Question 4

- Which of the formulas to the right is an inferential statistic calculating an estimate of the population standard deviation based on sample data?



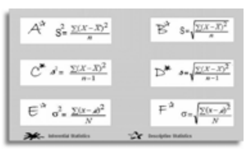
Question 4 Answer

- Formula "D" to the right is an inferential statistic calculating an estimate of the population standard deviation based on sample data?

D $$s = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$$

Question 5

- Which of the formulas to the right is a descriptive statistic calculating a standard deviation on sample data?



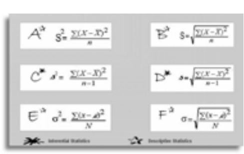
Question 5 Answer

- Formula "B" to the right is a descriptive statistic calculating a standard deviation on sample data?

B $$S = \sqrt{\frac{\sum(X-\bar{X})^2}{n}}$$

Question 6

- Which of the formulas to the right is a descriptive statistic calculating a variance on the actual population data?



Question 6 Answer

- Formula "E" to the right is a descriptive statistic calculating a variance on the actual population data?

E $$\sigma^2 = \frac{\sum(x-\mu)^2}{N}$$

Question 7

- Why is it that the estimated population variance and standard deviation based on sample data, inferential statistics (at left below), are always larger than the descriptive values calculated on the sample data(at right below)?

$$\hat{s}^2 = \frac{\sum(X-\bar{X})^2}{n-1} \qquad S^2 = \frac{\sum(X-\bar{X})^2}{n}$$

$$\hat{s} = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}} \qquad S = \sqrt{\frac{\sum(X-\bar{X})^2}{n}}$$

19

Question 7 Answer

- The values calculated at the left will always be larger due to the fact that the variation(SS) is divided by n-1(at left) rather than n(at right). This smaller value in the denominator then produces a larger calculated end value for the variance & thereafter a correspondingly larger value for the estimated standard deviation.

$$\hat{s}^2 = \frac{\sum(X-\bar{X})^2}{n-1} \qquad S^2 = \frac{\sum(X-\bar{X})^2}{n}$$

$$\hat{s} = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}} \qquad S = \sqrt{\frac{\sum(X-\bar{X})^2}{n}}$$

20

Question 8

- Why are the formulas for the inferential statistics at the left ADJUSTED by changing the n(at right) to n-1 in the formulas at the left?

$$\hat{s}^2 = \frac{\sum(X-\bar{X})^2}{n-1} \qquad S^2 = \frac{\sum(X-\bar{X})^2}{n}$$

$$\hat{s} = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}} \qquad S = \sqrt{\frac{\sum(X-\bar{X})^2}{n}}$$

21

Question 8 Answer

- The reason the inferential formulas at the left are made to be larger by using **n-1** rather than **n** is based on the realization that the formulas at the right underestimate the true values of these statistics in the population. To adjust for this underestimation, the inferential formulas are "corrected" via the **n-1** change to make the estimate more accurate.

$$\hat{s}^2 = \frac{\sum(X-\bar{X})^2}{n-1} \qquad S^2 = \frac{\sum(X-\bar{X})^2}{n}$$

$$\hat{s} = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}} \qquad S = \sqrt{\frac{\sum(X-\bar{X})^2}{n}}$$

22

Question 9

- How will you know if your data is POPULATION or SAMPLE data? and;
- When will you know that DESCRIPTIVE statistics versus INFERENTIAL statistics are called for?

23

Question 9 Answer

- You will know if your data is SAMPLE or POPULATION data on an exam because I will tell you what the data is.
- Furthermore, if you do not see any reference to a person's wanting to "estimate" anything, then what is being looked for are descriptive statistics. Wording referring to an interest in "describing" the data, either population or sample, invariably will be looking for "descriptive statistics."
- In the world outside the classroom, a researcher will know if her data is to be treated as a POPULATION, e.g., as I would when doing descriptive statistics on the EXAM-1 performance of my PSY 100 class(my population), and when it is to be treated as a SAMPLE. In an experiment, for example, most researchers work with SAMPLES but they are interested in estimating statistics for POPULATIONS.

24

Question 10

- What formulas can I use for calculating z-scores and what are their components?

25

Question 10 Answer

The formulas that you can use are shown below. Use those at the left if your data is sample data, those at the right if population data. In each case, a z-score is calculated by dividing a deviation from the mean value by a standard deviation.

Use with Sample Data

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{n}}$$

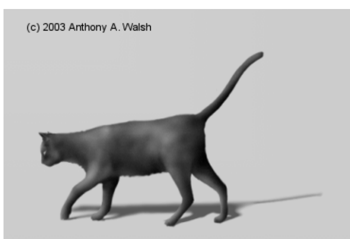
$$Z = \frac{X - \bar{X}}{S}$$

Use with Population Data

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

$$z = \frac{X - \mu}{\sigma}$$

26



27