PSY 3307

Variability

Jonathan A. Pedroza, MS, MA

Cal Poly Pomona

2021-08-31

Agenda

- Go over problem set 1
- Understanding What Variability is
- The Range
- Sample Variance
- Sample Standard Deviation
- Standard Deviation In a Normal Distribution
- Population Variance
- Population Standard Deviation
- Calculating Variance & Standard Deviation
    - Sample Variance
    - Sample Standard Deviation
    - Population Variance
    - Population Standard Deviation
- Statistics in Research (Means & Variability)

Problem Set 1

- Any questions regarding the assignment
- What parts need more clarification?
- Rating your own understanding of frequencies
- Rating your own understanding of measures of central tendency
- Who is ready to bring on the statistics?

A Little Review

```
numbers1 = c(7 , 6 , 3 , 3 , 1)
numbers2 = c(3 , 4 , 4 , 5 , 4)
numbers3 = c(4 , 4 , 4 , 4 , 4)

(7 + 6 + 3 + 3 + 1)/5
```

```
## [1] 4
```

```
(3 + 4 + 4 + 5 + 4)/5
```

```
## [1] 4
```

```
(4 + 4 + 4 + 4 + 4)/5
```

```
## [1] 4
```

```
number_ex <- tibble(numbers1, numbers2, numbers3)
```

- What is the mean of these three different sets of numbers?
- How are they different?

```r
psych::describe(number_ex)
```

```
##          vars n mean   sd median trimmed  mad min max range skew kurtosis
## numbers1    1 5    4 2.45      3       4 2.97   1   7     6 0.08     -2.0 1
## numbers2    2 5    4 0.71      4       4 0.00   3   5     2 0.00     -1.4 0
## numbers3    3 5    4 0.00      4       4 0.00   4   4     0  NaN      NaN 0
```

```r
numbers1 - 4.0
```

```
## [1]  3  2 -1 -1 -3
```

```r
numbers2 - 4.0
```

```
## [1] -1  0  0  1  0
```

```r
numbers3 - 4.0
```

```
## [1] 0 0 0 0 0
```

What are Measures of Variability?

- **Measures of Variability** are measures that examine how much scores differ from one another in a distribution
    - So if numbers are farther from one another, they will have more variability
    - It is also referred to as dispersion or spread; I refer to it as variation sometimes
- Low variability means that most values are close together, high variability means that values are spread out farther from one another
- Variability tells you how much your measure of central tendency is accurately measuring your distribution

Several Measures of Variability

- **Range**
- **Variance**
- **Standard Deviation**

The Range

$$Range = highest\ score - lowest\ score$$

- **Range** tells you the distance between your highest score/value and your lowest score/value
- The book states that the use of the range is mostly for nominal or ordinal data.
    - JP disagrees. The range or more importantly, the highest and lowest scores are very important for knowing how extreme some observations are
    - Also important for seeing right away that there are outliers

Sample Variance

- Variance tells you how different scores are from one another collectively
- This is measured by calculating how much all the scores vary/differ from the mean
- When we were looking at individual scores and how they differed from the mean, we looked at how much they deviated from the mean
- Now we are looking collectively at the sample
- We can think of variance as collective distance of the deviations from the mean

Some more review

$$X - \overline{X}$$

$$\Sigma(X - \overline{X})$$

- The Sum of deviations will always equal zero.

What is Sample Variance

- Because the sum will always equal zero, one way to get away from that is to square the sum of deviations

$$S_x^2 = \frac{\Sigma(X - \overline{X})^2}{N}$$

- **Sample Variance** is the average of the squared deviations of the scores around the sample mean
- To get the variance, we will calculate the sum of the deviations squared by the total observations.
- This gives us the distance scores are on average from the mean
- One issue is that the variance is a large number, which can get confusing at times
- Also, interpretation is difficult due to being squared

Sample Standard Deviation

- **Sample Standard Deviation** is the square root of the variance. It is a better measure of variability that better shows the average of the deviations from the mean

$$S_x = \sqrt{\frac{\Sigma(X - \overline{X})^2}{N}}$$

- Standard deviation and variance are very similar measures of variability
  - Both measure how much scores spread out from the mean
- Best interpretation is that the scores are "on average" this far from the mean

Standard Deviation and Area Under the Curve

- We can use the normal distribution/bell-shaped curve/normal curve to then add/subtract standard deviations from the mean
- If the average test score is 92 and the standard deviation is 2.5, then stating you are one standard deviation above the average score means you got a 94.5
- Going back to proportions and frequencies, knowing standard deviations means you can understand how much of a normal distribution is accounted for
- When you are looking at +1SD you have 34% of the distribution accounted for, when +-1SD you then account for 68% of the distribution

Population Standard Deviation & Variance

- **Population standard deviation** is the square root of the average squared deviation of scores around the population mean

$$\sigma_x = \sqrt{\frac{\Sigma(X-\mu)^2}{N}}$$

- **Population variance** the average squared deviation of scores around the population mean

$$\sigma_x^2 = \frac{\Sigma(X-\mu)^2}{N}$$

- These values are for the true population's measures of variability

Population Standard Deviation & Variance

- Examination of the population standard deviation and variance allows for us to interpret findings by using the normal distribution

Estimating the Population Variance & Standard Deviation

- The sample variance and sample standard deviations are only used to describe the variability of the sample
- They are **biased estimators**, which tend to underestimate the population variability by only using N
- We need a random sample to estimate a population, so we need a random sample of deviations
- We therefore need to get a sample of deviations, and N minus 1 reflects the variability of a population
- **Unbiased estimators** use sample data to estimate the population variability

Unbiased Estimators

$$s_x^2 = \frac{\Sigma(X - \overline{X})^2}{N - 1}$$

- **Estimated population variance** gives us the estimated amount of variation based on the sample we have

$$s_x = \sqrt{\frac{\Sigma(X - \overline{X})^2}{N - 1}}$$

- **Estimated Population Standard Deviation** is the standardized estimated amount of variation

Interpretation of Estimated Measures of Variability

- Interpretation is how we would expect the distribution to be.
- We might expect that 68% of the distribution is within the +-1SD of the mean

Summary

- Descriptive variance and standard deviation is for the sample
- When wanting to know how much scores vary from the population, we use the population variance and standard deviation
- When we are inferring about the population based on the sample we have, we compute the unbiased estimators, essentially meaning that we calculate N - 1 in our calculations

Computing the Formulas for Variance & Standard Deviation

- **Sum of Squared Xs** is calculating by squaring each score and then adding those values together

$$\Sigma X^2$$

- **Squared Sum of X** is calculated by adding all scores then squaring the sum

$$(\Sigma X)^2$$

- Sample Variance

$$S_x^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N}$$

Step 1: Find the sum of X

Step 2: Find the sum of squared X

Step 3: Square the value in step 1

Step 4: Divide the value in step 3 by N

Step 5: Subtract Step 2 from Step 4 values

Step 6: Divide by N

- Sample Standard Deviation

$$S_x = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N}}$$

Step 1: Find the sum of X

Step 2: Find the sum of squared X

Step 3: Square the value in step 1

Step 4: Divide the value in step 3 by N

Step 5: Subtract Step 2 from Step 4 values

Step 6: Divide by N

Step 7: get the square root of the value in Step 6

- Estimated Population Variance

$$s_x^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N - 1}$$

Step 1: Find the sum of X

Step 2: Find the sum of squared X

Step 3: Square the value in step 1

Step 4: Divide the value in step 3 by N

Step 5: Subtract Step 2 from Step 4 values

Step 6: Subtract 1 from N

Step 7: Divide by N

- Estimated Population Standard Deviation

$$s_x = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N-1}}$$

Step 1: Find the sum of X

Step 2: Find the sum of squared X

Step 3: Square the value in step 1

Step 4: Divide the value in step 3 by N

Step 5: Subtract Step 2 from Step 4 values

Step 6: Subtract 1 from N

Step 6: Divide by N

Step 7: get the square root of the value in Step 6

Calculating with the previous formulas

$$s_x^2 = \frac{\Sigma(X - \overline{X})^2}{N - 1}$$

Step 1: Get the mean

Step 2: Calculate the deviates from the mean

Step 3: Square the deviates

Step 4: Calculate the sum of the squared deviates

Step 5: Subtract 1 from N

Step 6: Divide numerator by the denominator

$$s_x = \sqrt{\frac{\Sigma(X - \overline{X})^2}{N - 1}}$$

Step 1: Get the mean

Step 2: Calculate the deviates from the mean

Step 3: Square the deviates

Step 4: Calculate the sum of the squared deviates

Step 5: Subtract 1 from N

Step 6: Divide numerator by the denominator

Step: Get the square root of the answer