homework_problem_set_2

JP

9/1/2021

1. Tell me what each of these symbols mean. Try this without looking at your notes.

$$\sigma$$

Answer sigma, which is the population standard deviation

$$\Sigma$$

Answer capital sigma, which here is the sum

$$\mu$$

Answer mu, which is the population mean

$$\overline{X}$$

Answer xbar, which is the sample mean

$$S \ or \ s \ or \ S_X \ or \ s_X$$

Answer standard deviation, which is the measure of variability for the sample or more likely to estimate the population

$$f$$

Answer frequency

$$N$$

Answer N for the total number of observations

2. If I were to give you the following scores/values, what would the standard deviation be?

10, 10

Answer 0 because there is no distance/dispersion/variability/variation between any of the scores it would be zero. If the scores don't vary, the sd will be zero

3. What is the range indirectly useful for finding?

Answer for showing the maximum and minimum values. More importantly for seeing if there were issues in coding your variables or for clearly visible outliers (those that don't take your survey seriously)

4. What is the difference between variance and standard deviation? What is similar about variance and standard deviation?

Answer They are both measures of variability. Variance is a larger value because it is measured in squared units while standard deviation is in easier to interpret units of whatever the variable is. For example, if it is age, the variation will be in days/months/years depending on how the variable is measured.

5. How do we get around the sum of deviations equaling zero? What are we doing to calculate for variance and standard deviation?

Answer The sum of deviations will always be zero unless you square the whole formula for the sum of deviations. Squaring it always for you to have all positive values which is helpful for calculating variance and standard deviation.

6. What is the area under the curve/proportion at or below the following z-scores?

```
##        z
## 1   1.40
## 2  -0.40
## 3   3.10
## 4   0.80
## 5  -1.70
## 6   2.57
## 7   1.67
## 8   1.11
## 9   0.00
## 10 -2.99
```

Answer For the following z-scores, the area under the curve is: 1.40 = .081 + .5 = .581 -.4 = .00003 3.1 = .999 .8 = .7881 -1.70 = .045 2.57 = .9949 1.67 = .953 1.11 = .867 0 = .5 -2.99 = .00139

7. What is the area between the mean and the following z-scores?

```
##        z
## 1   1.40
## 2  -0.40
## 3   3.10
## 4   0.80
## 5  -1.70
## 6   2.57
## 7   1.67
## 8   1.11
## 9   0.00
## 10 -2.99
```

Answer For the following z-scores, the area between the mean and these scores are: 1.40 = .081 -.4 = .49997 3.1 = .499 .8 = .2881 -1.70 = .455 2.57 = .4949 1.67 = .453 1.11 = .367 0 = 0 -2.99 = .49861

8. What is the percentile for each of the following z-scores?

```
##        z
## 1   1.40
## 2  -0.40
## 3   3.10
## 4   0.80
## 5  -1.70
## 6   2.57
## 7   1.67
## 8   1.11
## 9   0.00
## 10 -2.99
```

Answer For the following z-scores, the percentiles are: 1.40 = 58% -.4 = < 1% 3.1 = 99% .8 = 79% -1.70 = 4.5% 2.57 = 99% 1.67 = 95% 1.11 = 86.7% 0 = 50% -2.99 = < 1%

```
numbers
```

```
##  [1]  7.182374 10.378646  8.706081 11.553582  7.238934  9.792132 13.106835
##  [8]  9.868876 10.794828  4.422300 11.985773  8.801879  8.520575  9.290896
## [15]  7.579638
```

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

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

1. Calculate the variance and the standard deviation of the estimated population using the two formulas above.

```
7.182374+10.378646+8.706081+11.553582+7.238934+9.792132 + 13.106835 +9.868876+10.794828 +4.422300+11.98
```

```
## [1] 139.2233
```

```
# sum is 139.22

139.22/15
```

```
## [1] 9.281333
```

```
# mean is 9.28

# deviates
7.182374 - 9.28
```

```
## [1] -2.097626
```

```r
10.378646 - 9.28
```

```
## [1] 1.098646
```

```r
8.706081 - 9.28
```

```
## [1] -0.573919
```

```r
11.553582 - 9.28
```

```
## [1] 2.273582
```

```r
7.238934 - 9.28
```

```
## [1] -2.041066
```

```r
9.792132 - 9.28
```

```
## [1] 0.512132
```

```r
13.106835 - 9.28
```

```
## [1] 3.826835
```

```r
9.868876 - 9.28
```

```
## [1] 0.588876
```

```r
10.794828 - 9.28
```

```
## [1] 1.514828
```

```r
4.422300 - 9.28
```

```
## [1] -4.8577
```

```r
11.985773 - 9.28
```

```
## [1] 2.705773
```

```r
8.801879 - 9.28
```

```
## [1] -0.478121
```

```r
8.520575 - 9.28
```

```
## [1] -0.759425
```

```r
9.290896 - 9.28
```

```
## [1] 0.010896
```

```r
7.579638 - 9.28
```

```
## [1] -1.700362
```

```r
(-2.097626)^2
```

```
## [1] 4.400035
```

```r
(1.098646)^2
```

```
## [1] 1.207023
```

```r
(-0.573919)^2
```

```
## [1] 0.329383
```

```r
(2.273582)^2
```

```
## [1] 5.169175
```

```r
(-2.041066)^2
```

```
## [1] 4.16595
```

```r
(0.512132)^2
```

```
## [1] 0.2622792
```

```r
(3.826835)^2
```

```
## [1] 14.64467
```

```r
(0.588876)^2
```

```
## [1] 0.3467749
```

```r
(1.514828)^2
```

```
## [1] 2.294704
```

```r
(-4.8577)^2
```

```
## [1] 23.59725
```

```r
(2.705773)^2
```

```
## [1] 7.321208
```

```r
(-0.478121)^2
```

```
## [1] 0.2285997
```

```r
(-0.759425)^2
```

```
## [1] 0.5767263
```

```r
(0.010896)^2
```

```
## [1] 0.0001187228
```

```r
(-1.700362)^2
```

```
## [1] 2.891231
```

```r
4.400035 + 1.207023 +0.329383 + 5.169175 + 4.16595 + 0.2622792 + 14.64467 + 0.3467749 + 2.294704 + 23.5
```

```
## [1] 67.43513
```

```r
# numerator is 67.44
15 - 1
```

```
## [1] 14
```

```r
# denominator is 14
67.44/14
```

```
## [1] 4.817143
```

```r
# variance is 4.82
sqrt(4.82)
```

```
## [1] 2.19545
```

```r
# sd is 2.20


# double checked and it looks good
sd(numbers)
```

```
## [1] 2.194719
```

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

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

2. Use the two formulas above to calculate the variance and standard deviation.

```r
7.182374+10.378646+8.706081+11.553582+7.238934+9.792132 + 13.106835 +9.868876+10.794828 +4.422300+11.985
```

```
## [1] 139.2233
```

```r
# sum is 139.22

139.22^2
```

```
## [1] 19382.21
```

```r
# 19382.21 squared sum of x

19382.21/15
```

```
## [1] 1292.147
```

```r
# divied by N gives me 1292.147

7.182374^2+10.378646^2+8.706081^2+11.553582^2+7.238934^2+9.792132^2 + 13.106835^2 +9.868876^2+10.794828
```

```
## [1] 1359.644
```

```r
# sum of squared Xs is 1359.644

1359.644-1292.147
```

```
## [1] 67.497
```

```r
# numerator is 67.497

15-1
```

```
## [1] 14
```

```r
# denominator is 14

67.497/14
```

```
## [1] 4.821214
```

```r
# 4.82 is the variance

sqrt(4.82)
```

```
## [1] 2.19545
```

```r
# sd is 2.20
```

Range

```r
data_example <- data.frame(Maximum = c(30, 35, 40, 98, 47, 51, 61, 91, 41, 50),
                           Minimum = c(20, 29, 7, 64, 40, 34, 50, 14, 10, 12),
                           answers = c(10, 6, 33, 34, 7, 17, 11, 77, 31, 38))

data_example
```

```
##    Maximum Minimum answers
## 1       30      20      10
## 2       35      29       6
## 3       40       7      33
## 4       98      64      34
## 5       47      40       7
## 6       51      34      17
## 7       61      50      11
## 8       91      14      77
## 9       41      10      31
## 10      50      12      38
```

z-Score

$$z = \frac{X - \overline{X}}{S_X}$$

3. Calculate the z-score using the formula above

```r
get_z <- data.frame(observation = c(12, 14, 18, 5, 28 ,30, 14, 6, 14, 15),
                    mean = c(15, 16, 24, 7, 21, 24, 10, 8, 9, 4),
                    sd = c(3, 2, 4, 3.8, 1.4, 5.1, .9, .47, .21, .5))

get_z
```

```
##    observation mean   sd
## 1           12   15 3.00
## 2           14   16 2.00
## 3           18   24 4.00
## 4            5    7 3.80
## 5           28   21 1.40
## 6           30   24 5.10
## 7           14   10 0.90
## 8            6    8 0.47
## 9           14    9 0.21
## 10          15    4 0.50
```

```
z_fun <- function(observation, mean, sd){

  (get_z$observation - get_z$mean)/get_z$sd

}

z_fun(observation = observation, mean = mean, sd = sd)
```

```
##  [1] -1.0000000 -1.0000000 -1.5000000 -0.5263158  5.0000000  1.1764706
##  [7]  4.4444444 -4.2553191 23.8095238 22.0000000
```

```
# answers -1, -1, -1.5, -.53, 5, 1.18, 4.44, -4.26, 23.81, 22
```

Raw Score From z-Score

$$X = (z)(S_X) + \overline{X}$$

4. Calculate the raw score from the z-score using the formula above.

```
get_raw <- data.frame(z = c(1, 2, 3, -3, -2, -1, -3.7, 3.1, 1.5, 1.96),
                      mean = c(14, 16, 30, 13, 18, 19, 5, 16, 27, 20),
                      sd = c(2.1, 3.1, 1.4, 3, 1.2, .2, .9, .47, .16, .67))

get_raw
```

```
##        z mean   sd
## 1   1.00   14 2.10
## 2   2.00   16 3.10
## 3   3.00   30 1.40
## 4  -3.00   13 3.00
## 5  -2.00   18 1.20
## 6  -1.00   19 0.20
## 7  -3.70    5 0.90
## 8   3.10   16 0.47
## 9   1.50   27 0.16
## 10  1.96   20 0.67
```

```r
raw_fun <- function(z, mean, sd){
  get_raw$z*get_raw$sd + get_raw$mean
}


raw_fun(z = z, mean = mean, sd = sd)
```

```
##  [1] 16.1000 22.2000 34.2000  4.0000 15.6000 18.8000  1.6700 17.4570 27.2400
## [10] 21.3132
```

```r
# answers 16.1, 22.2, 34.2, 4, 15.6, 18.8, 1.67, 17.46, 27.24, 21.31
```

Standard Error of the Mean

$$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{N}}$$

5. Calculate the standard error of the mean using the formula above.

```r
get_sample_distribution <- data.frame(each_sample_mean = c(17, 15, 24, 61, 9, 10, 14, 18, 30, 4),
                                   mu = c(12, 16, 14, 22, 30, 40, 32, 54, 43, 8),
                                   sigma_x = c(3, 2, 4, 1.4, .2, .6, .7, .1, 2.4, 2.9), #this is not sig
                                   N = c(30, 15, 60, 17, 15, 34, 43, 60, 58, 81),
                                   se = c(.55, .52, .52, .34, .05, .10, .11, .01, .32, .32))

get_sample_distribution
```

```
##    each_sample_mean mu sigma_x  N   se
## 1                17 12     3.0 30 0.55
## 2                15 16     2.0 15 0.52
## 3                24 14     4.0 60 0.52
## 4                61 22     1.4 17 0.34
## 5                 9 30     0.2 15 0.05
## 6                10 40     0.6 34 0.10
## 7                14 32     0.7 43 0.11
## 8                18 54     0.1 60 0.01
## 9                30 43     2.4 58 0.32
## 10                4  8     2.9 81 0.32
```

```r
se_fun <- function(sigma_x, N){
  (get_sample_distribution$sigma_x)/sqrt(get_sample_distribution$N)
}

se_fun(sigma_x = sigma_x, N = N)
```

```
##  [1] 0.54772256 0.51639778 0.51639778 0.33954988 0.05163978 0.10289915
##  [7] 0.10674900 0.01290994 0.31513544 0.32222222
```

```r
# answers .55, .52, .52, .34, .05, .10, .11, .01, .32, .32
```

$$z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

```r
get_sample_distribution
```

```
##    each_sample_mean mu sigma_x  N   se
## 1                17 12     3.0 30 0.55
## 2                15 16     2.0 15 0.52
## 3                24 14     4.0 60 0.52
## 4                61 22     1.4 17 0.34
## 5                 9 30     0.2 15 0.05
## 6                10 40     0.6 34 0.10
## 7                14 32     0.7 43 0.11
## 8                18 54     0.1 60 0.01
## 9                30 43     2.4 58 0.32
## 10                4  8     2.9 81 0.32
```

```r
z_se_fun <- function(each_sample_mean, mu, se){
  (get_sample_distribution$each_sample_mean - get_sample_distribution$mu)/get_sample_distribution$e
}

z_se_fun(each_sample_mean = each_sample_mean, mu = mu, se = se)
```

```
##  [1]  0.29411765 -0.06666667  0.41666667  0.63934426 -2.33333333 -3.00000000
##  [7] -1.28571429 -2.00000000 -0.43333333 -1.00000000
```

```r
# answers
# .29, -.067, .417, .64, -2.33, -3, -1.29, -2, -.43, -1
```

6. Using the formula above, calculate the z score from the standard error of the mean (**Note** remember that you will be using the standard deviation of the sample mean and not the standard deviation of each sample.)

7. Calculate the estimated population variance and standard deviation using either the formulas used for question 1 or the formulas used for question 2.

```r
numbers2
```

```
##  [1] 5.060850 3.865486 3.521444 3.571518 4.629313 4.115526 4.492806 4.364263
##  [9] 5.067677 2.977059 3.984693 6.151861 3.580645 2.844179 4.676546
```

```r
5.060850+3.865486+3.521444+3.571518+4.629313+4.115526+4.492806+4.364263+5.067677+2.977059+3.984693+6.15
```

```
## [1] 62.90387
```

```r
# 62.90 sum of numbers
62.90^2
```

```
## [1] 3956.41
```

```r
# 3956.41 squared sum of x
3956.41/15
```

```
## [1] 263.7607
```

```r
# divided by N gives me 263.7607
5.060850^2+3.865486^2+3.521444^2+3.571518^2+4.629313^2+4.115526^2+4.492806^2+4.364263^2+5.067677^2+2.97
6.151861^2+3.580645^2+2.844179^2+4.676546^2
```

```
## [1] 274.3585
```

```r
# sum of squared Xs is 274.36
274.36 - 263.76
```

```
## [1] 10.6
```

```r
# numerator is 10.6
15-1
```

```
## [1] 14
```

```r
# denominator is 14
10.6/14
```

```
## [1] 0.7571429
```

```r
# .76 is the variance
sqrt(.76)
```

```
## [1] 0.8717798
```

```r
# sd is .87
sd(numbers2)
```

```
## [1] 0.8687202
```