**ITEC 610**

**Applied Managerial Statistics**

**Describing Data Assignment**

Show all work.

Unsubstantiated answers receive no credit.

Be sure to attach all output from Excel/R/R.

No deadline extension will be given.

Name: Stefano Musetti-Suarez

Honor Code

All work on this assignment is my own. I have not discussed this assignment with anyone (except for Dr. Mahsa Oroojeni) in any way (including, but not limited to, text messages, email, face-to-face, fax, tweets, etc.) Signature: Stefano Musetti-Suarez

**Describing Data**

1) The following is a list of prices (in dollars) of birthday cards found in various drug stores:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 2.45 | 1.20 | 0.85 | 1.33 | 2.25 |
| 2.25 | 2.09 | 2.99 | 1.00 | 0.88 |
| 1.42 | 2.36 | 2.15 | 2.85 | 1.52 |
| 1.99 | 2.38 | 0.85 | 2.22 | 2.75 |

1. *Using Excel/R find the mean, median, mode, range, variance, and standard deviation of the data. Attach your output from Excel/R.*

**=** (2.45 + 2.25+ . . . + 2.75)/20 = **$1.889**

**Median**= $0.85 . . . $2.09 + $2.15/2 . . . $2.99 (n = 20)

= (**x10 + x11/2)** = (2.09 + 2.15)/2 = **$2.12**

**Mode** = (More than 1 mode, two numbers repeat 2 times = **$0.85 & $2.25**)

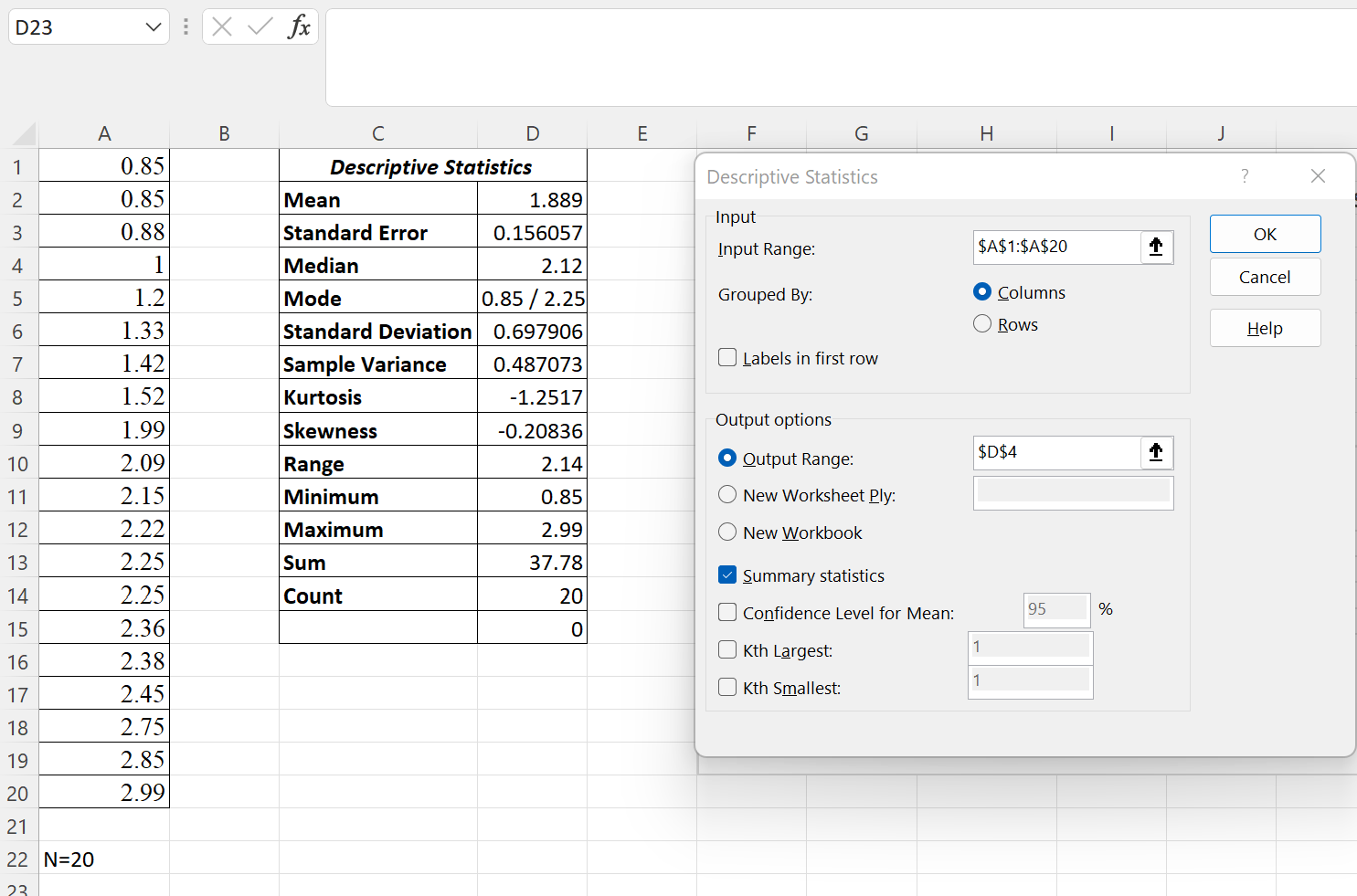
**Range** = (xlargest – xsmallest)

= ($2.99 - $0.85) = $**2.14**

²= [(2.45 – 1.889)2 + (2.25 – 1.889)2 +...+ (2.75 -1.889)2] / (20-1)

**= $0.4870**

**S = = $0.6979**

**Output from Excel:**

|  |  |
| --- | --- |
| ***Column1*** |  |
| **Mean** | 1.889 |
| **Standard Error** | 0.156057 |
| **Median** | 2.12 |
| **Mode** | 0.85 / 2.25 |
| **Standard Deviation** | 0.697906 |
| **Sample Variance** | 0.487073 |
| **Kurtosis** | -1.2517 |
| **Skewness** | -0.20836 |
| **Range** | 2.14 |
| **Minimum** | 0.85 |
| **Maximum** | 2.99 |
| **Sum** | 37.78 |
| **Count** | 20 |
|  | 0 |

Output from R:

Graphical user interface, application

Description automatically generated

1. *Using Excel/R, construct a frequency histogram of the data set. Use the guidelines in the class notes. Provide all the details (interval, width, etc.) on how you constructed the histogram. Make sure that you attach the histogram created by Excel/R. Comment on the shape of the frequency distribution (e.g., is the distribution skewed? Is the distribution approximately mound-shaped and symmetric?) for the data set based on your histogram.*

**Intervals**= = = **4.4721 = 5 bins when rounding up.**

**Width=** (xlargest – xsmallest)/5 = ($2.99 - $0.85)/5 = **0.428 = 0.5 (rounding up).**

* Started the first interval -0.05 from the minimum valor (0.85-0.05) Later, each bin respects the width of 0.5.

|  |  |
| --- | --- |
| *Bin* | *Frequency* |
| 0.8 | 0 |
| 1.3 | 5 |
| 1.8 | 3 |
| 2.3 | 6 |
| 2.8 | 4 |
| 3.3 | 2 |
| More | 0 |

**Graphical user interface, application

Description automatically generated**

**Text

Description automatically generatedChart, histogram

Description automatically generatedR Output:**

* As the mean is smaller than the median, the implication is that the distribution is skewed to the left and therefore there is no normal distribution. This can be noted in the histogram because there is no bell-shaped curve, and the values are not evenly distributed near the mean.

1. *Based on the results in part a, construct the intervals and for the data set. Be sure to show your interval below. Based on the results in part b what percentage of the measurements for the data set falls in each interval?*



**1 SD**: – s to + s = 1.889 – 0.6979=**1.1911**; 1.889 + 0.6979=**2.5869**

**2 SD**: – 2s to + 2s = 1.889 – 2(0.6979) = **0.4932;** 1.889 + 1.3958 = **3.2848**

**3 SD**: – 3s to + 3s = 1.889 – 3(0.6979) = **-0.2047**; 1.889 + 2.0937= **3.9827**

**1st Interval:** 13/20= 65%

**2nd Interval:** 20/20=100%

**3rd Interval:** 20/20=100%

* Chebyshev theorem provides an approximation of the datasets, but not exact values compared to the categorical empirical rule. Therefore, we can see that **65% of the datasets fall within the first interval** (which is not really useful with CH theorem). **100% of the datasets fall within the second interval**, which is something that could happened because the rule states that at least 75% of them must fall in this category. **Finally, 100% of the data points also fall within the last interval**, which is consistent with the theorem which states that 88.89% of the data points must fall within the third category.

2) *A semiconductor manufacturer produces printed circuit boards that are sampled to determine the thickness of their copper plating. The following statements create a data set named Trans, which contains the plating thicknesses (Thick) of 50 boards:*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3.412 | 3.45 | 3.551 | 3.451 | 3.60 | 3.462 | 3.586 | 3.645 | 3.252 | 3.62 |
| 3.606 | 3.634 | 3.852 | 3.56 | 3.342 | 3.341 | 3.444 | 3.774 | 3.632 | 3.199 |
| 3.71 | 3.654 | 3.723 | 3.981 | 3.934 | 3.708 | 3.934 | 3.315 | 3.762 | 3.223 |
| 3.469 | 3.481 | 3.515 | 3.535 | 3.46 | 3.575 | 3.488 | 3.515 | 3.484 | 3.482 |
| 3.517 | 3.483 | 3.467 | 3.467 | 3.502 | 3.471 | 3.516 | 3.474 | 3.5 | 3.466 |

1. Using Excel/R find the mean, median, mode, range, variance, and standard deviation of the data. Attach your output from Excel/R.

**=** (3.412 + 3.606+ . . . + 3.466)/50 = **3.5444**

**Median=** 3.199 . . . 3.5+ 3.502/2 . . . 3.981 **(n=50)**

**=** **(x25 + x26/2)** = (3.5 + 3.502)/2 = **3.5085**

**Sort Values … x10 + x11/2… $2.12**

**Mode** **=** (Multiple Modes, three numbers are repeated 2 times = **(3.467; 3.525; 3.934)**

**Range** **=** (xlargest – xsmallest)

**=** (3.981 - 3.199) = **0.782**

**S²** **=** [(3.412– 3.5444)2 + (3.606– 3.5444)2 +...+ (3.466-3.5444)2] / (50-1)

**= 0.028375193**

**S** = = **0.168449378152607**

Data from Excel

A picture containing graphical user interface

Description automatically generatedTable

Description automatically generated

Output from R:

Graphical user interface, text, application

Description automatically generated

1. Based on the results in part a, construct the intervals and for the data set. Be sure to show your interval below. What percentage of the measurements for the data set falls in each interval? Compare the intervals , and . Explain why the results are different.



**1 SD**: – s to + s = 3.5444– 0.1684 = **3.376031**; 3.5444+ 0.1684 = **3.712929**

**2 SD**: – 2s to + 2s = 3.5444– 2(0.1684) = **3.207581**; 3.5444+ 2(0.1684) = **3.881379**

**3 SD**: – 3s to + 3s = 3.5444– 3(0.1684) = **3.039132**; 3.5444+ 3(0.1684) = **4.049828**

**1st Interval:** 37/50= 65%

**2nd Interval:** 46/50=92%

**3rd Interval:** 50/50=100%

* Confidence Intervals portray the likely range of values of the population mean. In the first interval we can see an approximation of the 65% of the total of the data points, while in the second and third intervals we can see over 90% and 100% of the data points.