

**KULLIYAH OF INFORMATION AND COMMUNICATION TECHNOLOGY**

**CSC 1706 PROBABILITY AND STATISTICS**

**SEMESTER 2, 2017/2018**

**SECTION 4**

**GROUP 4**

**ASSIGNMENT 2: QUESTION 20**

**PREPARED BY:**

|  |  |
| --- | --- |
| **NAME** | **MATRIC NO.** |
| Ahmad Zaidan Bin Adnan | 1718733 |
| Muhammad Zaim Taqiyuddin Bin Zubir | 1714113 |
| Safiyyah Binti Ghazali Othman Taufik | 1718532 |
| Ahmad Zahin Fikri Bin Rozlee | 1713089 |
| Shafiq Azhar Bin Shahrizal | 1616313 |

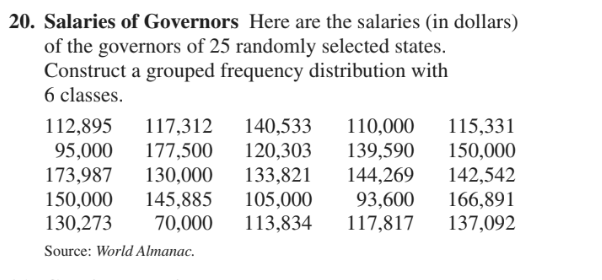
**LECTURER**

DR. AFIDALINA TUMIAN

**DUE**

25 MARCH 2018

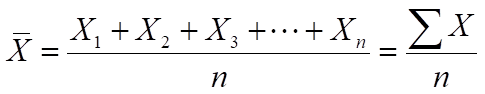
**QUESTION:**



**Measures of Central Tendency**

**Data Description**

To find the mean, the following formula is used:



The mean of the sample is 129339.

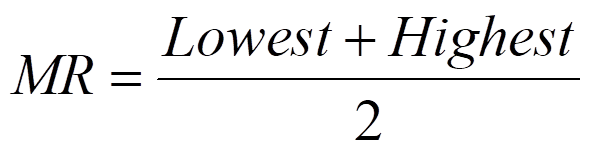
To find the median, arrange the data values in ascending order.

70000 93600 95000 105000 110000 112895 113834 115331 117312 117817 120303 130000 130273 133821 137092 139590 140533 142542 144269 145885 150000 150000 166891 173987 177500

Since the 13th number is the median, then the median is 130273.

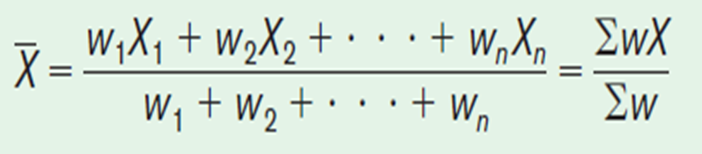
From the data, the mode is 150000 because the value occurs the most times which is two times.

To find the midrange, the following formula is used:



The midrange is 123750.

To find the weighted mean, the following formula is used:



The weighted mean is 129339.

**R Codes**

>mean(data)

[1] 129339

>median(data)

[1] 130273

>table.data &lt;- table(data)

>table.data[which(table.data == max(table.data)

[1] 150000

>(min(data) + max(data))/2

[1] 123750

>weighted.mean(data)

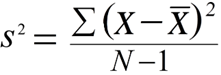
[1] 129339

**Measures of Variation**

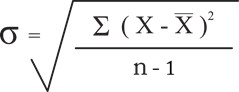
**Data Description**

Range is 70000 to 177500

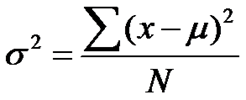
Variance

 = 648739766

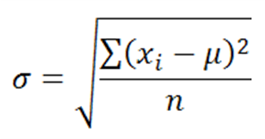
Standard deviation

 = 25470.37

Population variance

 =622790175

Population standard deviation

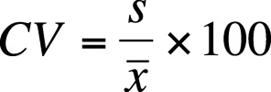


= 657.76 1373.56 1785.92 826.44 37.36 481.08 1926.44 26.44 661.84

2373.56 447.76 361.44 179.28 973.56 620.20 773.56 410.04 597.20

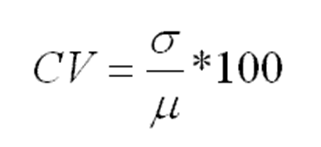
1429.56 460.88 560.32 826.44 528.12 1502.08 310.12

Coefficient of variance (for samples)



= 19.69272

Coefficient of variance (for population)



= 0.50855504 1.06198440 1.38080548 0.63897200 0.02888533 0.37195278 1.48945020 0.02044240 0.51170954 1.83514640 0.34619102 0.27945167 0.13861248 0.75271960 0.47951507 0.59808720 0.31702735 0.46173235 1.10528147 0.35633490 0.43321813 0.63897200 0.40832232 1.16135118 0.23977300

**R Codes**

> range(data)

[1] 70000 177500

> var(data)

[1] 648739766

> sd(data)

[1] 25470.37

> sum((data-mean(data))^2)/length(data)

[1] 622790175

> sqrt((data-mean(data))^2)/length(data)

[1] 657.76 1373.56 1785.92 826.44 37.36 481.08 1926.44 26.44 661.84

[10] 2373.56 447.76 361.44 179.28 973.56 620.20 773.56 410.04 597.20

[19] 1429.56 460.88 560.32 826.44 528.12 1502.08 310.12

> sd(data)\*100/mean(data)

[1] 19.69272

> sqrt((data-mean(data))^2)/length(data)\*100/mean(data)

[1] 0.50855504 1.06198440 1.38080548 0.63897200 0.02888533 0.37195278

[7] 1.48945020 0.02044240 0.51170954 1.83514640 0.34619102 0.27945167

[13] 0.13861248 0.75271960 0.47951507 0.59808720 0.31702735 0.46173235

[19] 1.10528147 0.35633490 0.43321813 0.63897200 0.40832232 1.16135118

[25] 0.23977300

**Measures of Position**

**Data Description**

Data sorting in increasing manner

Based on the data that we have, we need to sort the data in ascending order. In total, we

have 25 salaries of governors. We have to determine the lowest salary and the highest

salary to arrange it in sequence.

70000, 93600, 95000, 105000, 110000, 112895, 113834, 115331, 117312, 117817,

120303, 130000, 130273, 133821, 137092, 139590, 140533, 142542, 144269, 145885,

150000, 150000, 166891, 173987, 177500.

Standard scores ( or z-score )

The z score or standard score value is obtained by subtracting the mean from the value and

dividing the result by the standard deviation.

Mean = 129339

Standard deviation = 25470.37

The formula is :

Z = (value – mean) / standard deviation

Z = ( X – 129339 ) / 25470.37

Quartile

To get quartiles, we need to separate the data set into 4 equal groups.

First step : arrange the data :

70000, 93600, 95000, 105000, 110000, 112895, 113834, 115331, 117312, 117817,

120303, 130000, 130273, 133821, 137092, 139590, 140533, 142542, 144269, 145885,

150000, 150000, 166891, 173987, 177500.

Second step : median of the data value = Q2 = 130373

Third step : median of the data value fall below Q2 = Q1 = 113834

Forth step : median of the data value fall above Q2 = Q3 = 144269

Deciles

To get deciles, we need to separate the data set into 10 equal groups.

Percentile

Percentile is position in hundredths that a data value holds in the distribution.

**R Codes**

> sort(data, decreasing = FALSE)

[1] 70000 93600 95000 105000 110000 112895 113834 115331 117312 117817

[11] 120303 130000 130273 133821 137092 139590 140533 142542 144269 145885

[21] 150000 150000 166891 173987 177500

> summary (data)

Min. 1st Qu. Median Mean 3rd Qu. Max.

70000 113834 130273 129339 144269 177500

> (data - mean(data))/ sd(data)

[1] -0.64561291 -0.47219573 0.43949106 -0.75927439 -0.54997237 -1.34819398

[7] 1.89086375 -0.35476516 0.40246765 0.81117784 1.75293878 0.02595172

[13] 0.17596917 0.58617130 0.51836702 0.81117784 0.64961757 -0.95558092

[19] -1.40315981 1.47434056 0.03667006 -2.32972663 -0.60874655 -0.45236877

[25] 0.30439290

> quantile(data)

0% 25% 50% 75% 100%

70000 113834 130273 144269 177500

>quantile(data, prob = seq(0, 1, length = 11))

0% 10% 20% 30% 40% 50% 60% 70%

70000.0 99000.0 112316.0 115727.2 119308.6 130273.0 138091.2 142140.2

80% 90% 100%

146708.0 160134.6 177500.0

> quantile(data, prob = seq(0, 1, length = 101))

0% 1% 2% 3% 4% 5% 6% 7%

70000.0 75664.0 81328.0 86992.0 92656.0 93880.0 94216.0 94552.0

8% 9% 10% 11% 12% 13% 14% 15%

94888.0 96600.0 99000.0 101400.0 103800.0 105600.0 106800.0 108000.0

16% 17% 18% 19% 20% 21% 22% 23%

109200.0 110231.6 110926.4 111621.2 112316.0 112932.6 113157.9 113383.3

24% 25% 26% 27% 28% 29% 30% 31%

113608.6 113834.0 114193.3 114552.6 114911.8 115271.1 115727.2 116202.6

32% 33% 34% 35% 36% 37% 38% 39%

116678.1 117153.5 117392.8 117514.0 117635.2 117756.4 118115.3 118712.0

40% 41% 42% 43% 44% 45% 46% 47%

119308.6 119905.2 121078.8 123406.0 125733.3 128060.6 130010.9 130076.4

48% 49% 50% 51% 52% 53% 54% 55%

130142.0 130207.5 130273.0 131124.5 131976.0 132827.6 133679.1 134475.2

56% 57% 58% 59% 60% 61% 62% 63%

135260.2 136045.3 136830.3 137491.7 138091.2 138690.7 139290.2 139703.2

64% 65% 66% 67% 68% 69% 70% 71%

139929.5 140155.8 140382.1 140693.7 141175.9 141658.0 142140.2 142611.1

72% 73% 74% 75% 76% 77% 78% 79%

143025.6 143440.0 143854.5 144269.0 144656.8 145044.7 145432.5 145820.4

80% 81% 82% 83% 84% 85% 86% 87%

146708.0 147695.6 148683.2 149670.8 150000.0 150000.0 150000.0 150000.0

88% 89% 90% 91% 92% 93% 94% 95%

152026.9 156080.8 160134.6 164188.4 167458.7 169161.7 170864.8 172567.8

96% 97% 98% 99% 100%

174127.5 174970.6 175813.8 176656.9 177500.0

**Exploratory Data Analysis**

**Data Description**

The Five-Number Summary is composed of the following numbers: Low, Q1, MD, Q3, High.

The Five-Number Summary can be graphically represented using a Boxplot.

**R Codes**

> fivenum(data)

[1] 70000 113834 130273 144269 177500

> boxplot(data)

