Sample space

45, 80, 83, 78, 75, 77, 79, 83, 83, 100

Sorted vector: 45 75 77 78 79 80 83 83 83 100

1. (wrong, sort first, then take I5 + I6 /2 )Since sample length is even, then **median** is sum(i5+i6)/2 = 79+80/2 = **79.5,**
2. **mean = (sum/length) = (45+80+83+78+75+77+79+83+83+100)/10 = 78.3**

|  |  |  |
| --- | --- | --- |
| Xi | Xi-X’ | (Xi-x’)2 |
| 45 | -33.3 | 1108.89 |
| 80 | 1.7 | 2.89 |
| 83 | 4.7 | 22.09 |
| 78 | -0.3 | 0.09 |
| 75 | -3.3 | 10.89 |
| 77 | -1.3 | 1.69 |
| 79 | 0.7 | 0.49 |
| 83 | 4.7 | 22.09 |
| 83 | 4.7 | 22.09 |
| 100 | 21.7 | 470.89 |

S2 = 1662.1 / 9 = **184.67 = Variance**

S = **13.58 = Standard Deviation**

Q1 = element at index #3 = **77 (wrong, order first) R: quantile(data, 0.25)**

Q3 = element at index #8 = **83 (wrong, order first) R: quantile(data, 0.75)**

**Q2 = R: quantile(data, 0.5)**

Interquartile range (IRQ) = **6**

2

1. What is the probability that an individual has diabetes? (590/5000) \* 100 = **11.8%**
2. What is the probability that a randomly selected patient is an American Indian who is not a diabetic? (840/5000) \* 100 = **16.8%**
3. What is the probability that a randomly selected patient is an American Indian or a non diabetic?

P(A v B) = P(A) + P(b) – P(A|B) = 1000/5000 + 4400/5000 – 840/5000 = 0.912 \* 100 = 91.2

(4410 + 160)/5000 \* 100 = **91.4%**

1. What is the probability of being diabetic among Hispanics? (130 / 1000) \* 100 = **13%**