

Answers

| Question | Answer |
|----------|--------|
| 1 | D |
| 2 | C |
| 3 | A |
| 4 | B |
| 5 | B |
| 6 | C |
| 7 | D |
| 8 | A |
| 9 | D |
| 10 | C |
| 11 | B |
| 12 | D |
| 13 | A |
| 14 | D |
| 15 | C |
| 16 | C |

Q17, Velocity, Variability, Variety, Veracity

Q18. Predictive, Diagnostic, Embedded, Prescriptive

Q19.

- a. iii, iv
- b. iv
- c. Left-skewed. Average is affected by extreme values (small values in this case) and median is the 50th quantile, median is larger than mean.
- d. "i, ii" or "i, ii, iv" (since question did not state the purpose of the clustering analysis, year is a "valid" variable to use although not logical)

Q20.

- a. [Free form answers]
b.

| | |
|---|--|
| Variable(s) requiring (one-hot or dummy) encoding | Gender |
| Does variables standardization? | Yes. Reason: _The variables are in wide ranges _ (note: different ranges isn't good enough) |
| Target variable (select one or N/A if not applicable) | N/A |
| Independent variables (list all applicable ones) | All Age, weight, gender (M/F), average heart rate, amount of sleep in a week, average number of steps per week. |

Q21.

- a. 0.80 / 8:2 / 80:20 / 80% / 80%:20% (any other forms) 7.62
b. 0.65 5.54
c. F F T T
d. 1.17 1.67
e. T F T F F T
f. 0.009
g. 5 6
h. 0.87
i. [free form answers]