Baskent University - Computer Engineering
2024-2025 /Spring
Assist. Prof. Elmas Burcu MAMAK EKİNCİ 20.03.2025

BİL/CSE 395 APPLIED DATA ANALYSIS

QUIZ #1

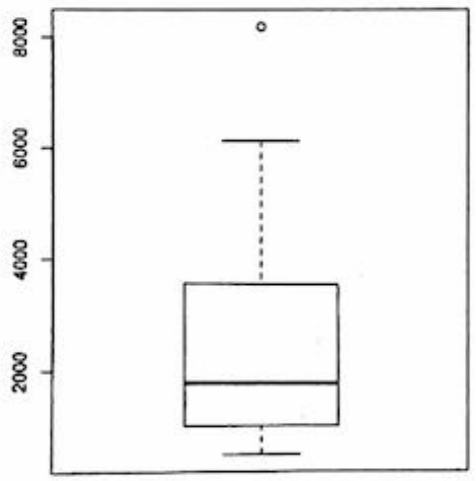
Name-Surname:.....# KEY #..... Number:.....

Question 1: Write the variable types and measurement levels of the variables given below.

Variables	Types of Variables (Quantitative/Qualitative)	Levels of Measurement (Nominal/ Ordinal/ Interval/ Ratio)
CPU Temperature (°C or °F)	(Nicel) Quantitative	Interval
Uptime of a Server (Hours)	(Nicel) Quantitative	Ratio
Number of Processors in a System (e.g., 1, 2, 4, 8)	(Nicel) Quantitative	Ratio
Network Protocol (TCP, UDP, HTTP, FTP)	Qualitative (Nitel)	Nominal
Internet Connection Speed Category (Slow, Moderate, Fast)	Qualitative (Nitel)	Ordinal

Question 2. A software engineering team collects memory usage (in MB) from different applications running on a server. They performed some analyses in the R program using this data. The functions and outputs used in R are listed below.

```
> memory_usage<-c(512, 1024,1024, 2048, 4096, 8192, 1024, 1536, 307
2, 6144, 2048,1536)
> summary(memory_usage)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    512    1024    1792    2688    3328    8192
> sd(memory_usage)
[1] 2352.618
> var(memory_usage)
[1] 5534813
> boxplot(memory_usage)
```



Baskent University - Computer Engineering
2024-2025 /Spring
Assist. Prof. Elmas Burcu MAMAK EKİNCİ

20.03.2025

BİL/CSE 395 APPLIED DATA ANALYSIS

QUIZ #1

a. What are the mean and median? What can you infer about the central tendency of memory usage from the mean and the median?
b. What do you say about the shape of the distribution? Why?
c. Based on the descriptive statistics, which data point do you identify as an outlier? How would removing the outlier from the dataset likely impact the summary statistics?

(a) Mean = $\bar{x}$ = 2688
Median = 1792
Since the mean is higher than the median, this suggests that there is positive/right skewness in the data. Meaning there are a few higher values pulling the mean up.

Min 512 — Max 8192

(b) Mean > Median and according to box plot ⇒ The data is right skewed.

(c) 8192 MB is an outlier. When looking at the box plot, it is seen that the maximum value is the outlier.
* Removing this outlier substantially lowers the mean and standard deviation, bringing the mean closer to the median and reducing the right skewed of distribution.

Question 3. A network engineering team is testing 20 network switches. The probability of a failure on any given switch is 0.15. Some R outputs for this problem are given below.

```
> dbinom(x=5,size=20, prob=0.15)
[1] 0.1028452
> dbinom(x=0,size=20, prob=0.15)
[1] 0.03875953
> dbinom(x=1,size=20, prob=0.15)
[1] 0.1367983
> pbinom(6,20,0.15) => P(X≤6) = F(6)
[1] 0.9780649
```

X: # of failures in 20 network switches.
$X \sim Binomial(p=0.15, n=20)$

a. What is the probability of having exactly 5 failures in 20 network switches?
b. What is the probability of having more than 6 failures in 20 network switches?
c. What is the probability of having 1 or fewer failures in 20 network switches?

(a) $P(X=5) = f(5) = 0.1028452$

(b) $P(X>6) = 1 - P(X \leq 6) = 1 - F(6) = 1 - 0.9780649 = 0.0219351$

(c) $P(X \leq 1) = P(X=0) + P(X=1) = f(0) + f(1) =$
$= 0.03875953 + 0.1367983$
$= 0.1755578$ //