| | |
|---|---|
| **Started on** | Monday, November 15, 2021, 9:40 AM |
| **State** | Finished |
| **Completed on** | Monday, November 15, 2021, 9:59 AM |
| **Time taken** | 18 mins 50 secs |
| **Grade** | **9.00** out of 10.00 (**90**%) |

Question **1**

Correct

1.00 points out of 1.00

We have a dataset as follows. Please answer the next 2 questions according to this dataset.

```
df.head(5)
```

| | carat | cut | color | clarity | depth | table | price | x | y | z |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.23 | Ideal | E | SI2 | 61.5 | 55.0 | 326 | 3.95 | 3.98 | 2.43 |
| 1 | 0.21 | Premium | E | SI1 | 59.8 | 61.0 | 326 | 3.89 | 3.84 | 2.31 |
| 2 | 0.23 | Good | E | VS1 | 56.9 | 65.0 | 327 | 4.05 | 4.07 | 2.31 |
| 3 | 0.29 | Premium | I | VS2 | 62.4 | 58.0 | 334 | 4.20 | 4.23 | 2.63 |
| 4 | 0.31 | Good | J | SI2 | 63.3 | 58.0 | 335 | 4.34 | 4.35 | 2.75 |

Which one of those regarding df.describe() function is correct?

a. It will show count, mean, std, min and max values and quartiles for all of the data, regardless of its type.

b. It will only show descriptive statistics for numerical data and only shows counts for categorical variables.

c. It will only show descriptive statistics for numerical data and bring nothing about categorical data in our dataset. ✔

d. It will only show descriptive statistics for categorical data, not the numerical data.

The correct answer is:
It will only show descriptive statistics for numerical data and bring nothing about categorical data in our dataset.

Question **2**

Correct

0.33 points out of 0.33

Mean, median and mode can only be calculated for categorical variables. For discrete and categorical variables, we generally use value counts.

Select one:

True

False ✔

The correct answer is 'False'.

Question **3**

Correct

1.00 points out of 1.00

Selen received 80 out of 100 from the statistics course (mean of the grades in the class is 30 and the standard deviation is 10). She also won 2 prizes out of 10 at the school's various competitions. Melek received 30 from the same statistics course and won 5 prizes in total from the competitions. If you are asked to compare two students' achievements at school, which of the following measures and metrics will be helpful to use? You can select more than one.

a. Risk ratio

b. Geometric mean ✔

c. Coefficient of variation

d. None of theö

The correct answer is:
Geometric mean

Question **4**

Correct

1.00 points out of 1.00

Regarding statistical measures, which one of those statements are correct? (Select all that apply).

- [x] a. We have a very small dataset where the mean, mode and median are equal to each other. If we add a new data instance the value of which is equal to the mean, the standard deviation of the dataset will not change. ✖
- [ ] b. When we compare two datasets with the same mean, the one with smaller standard deviation have a flatter normal curve.
- [ ] c. In a dataset, as we divide sample st. dev by (N-1) and population st. dev by N, we always have a higher standard deviation for any sample st. deviation.
- [x] d. A dataset can have more than one mode but cannot have more than one median. ✔
- [ ] e. If the dataset's mode is higher than median and median is higher than mean, this dataset have positive skewness.

The correct answer is:
A dataset can have more than one mode but cannot have more than one median.

Question **5**

Correct

1.00 points out of 1.00

Regarding pyplot and seaborn libraries, which one of those statements are correct? (Select all that apply).

- [x] a. In seaborn, we can show different levels of a categorical variable by passing "hue" parameter. ✔
- [ ] b. They add up to one actual library. For example, seaborn includes functions for drawing a violinspot but not for drawing histograms whereas pyplot have functions for drawing histograms but not for violinplots.
- [ ] c. plt.show() function is necessary to get be able to see the visual in Jupyter Notebook.
- [x] d. We can create our own palettes for color maps in seaborn library. ✔
- [x] e. You can visualize different types of plots in one subplot. ✔

The correct answers are:
You can visualize different types of plots in one subplot.,
We can create our own palettes for color maps in seaborn library.,
In seaborn, we can show different levels of a categorical variable by passing "hue" parameter.

Question **6**

Incorrect

0.00 points out of 1.00

By looking at the descriptive statistics below, when we draw the histogram of the variables, which one of those is the least likely to have a skewed distribution?

| | carat | price | x | y |
|---|---|---|---|---|
| count | 53940.000000 | 53940.000000 | 53940.000000 | 53940.000000 |
| mean | 0.797940 | 3932.799722 | 5.731157 | 5.734526 |
| std | 0.474011 | 3989.439738 | 1.121761 | 1.142135 |
| min | 0.200000 | 326.000000 | 0.000000 | 0.000000 |
| 25% | 0.400000 | 950.000000 | 4.710000 | 4.720000 |
| 50% | 0.700000 | 2401.000000 | 5.700000 | 5.710000 |
| 75% | 1.040000 | 5324.250000 | 6.540000 | 6.540000 |
| max | 5.010000 | 18823.000000 | 10.740000 | 58.900000 |

- ( ) a. Y
- ( ) b. X
- ( ) c. Carat
- (•) d. Price ✖

The correct answer is:
X

Question **7**

Correct

0.34 points out of 0.34

A Z score of 2.0 indicates that the value is located two standard deviations to the right of the mean.

Select one:

- (•) True ✔
- ( ) False

The correct answer is 'True'.

Question **8**

Correct

0.33 points out of 0.33

Among mean, mode and median, mean is the one who is affected by the extreme values the most.

Select one:

- True ✔
- False

The correct answer is 'True'.

Question **9**

Correct

1.00 points out of 1.00

Select the best central tendency measure for describing the data.

| | |
|---|---|
| Income of people in İstanbul where we have a lot of outliers. | Median ✔ |
| Attendance sheet (everyone is either present or not present) | Mode ✔ |
| Height of the population, where we have continuous and almost normally distributed data. | Mean ✔ |

The correct answer is: Income of people in İstanbul where we have a lot of outliers. → Median,
Attendance sheet (everyone is either present or not present) → Mode,
Height of the population, where we have continuous and almost normally distributed data. → Mean

Question **10**

Correct

1.00 points out of 1.00

Please select the correct data type for each of the case.

| | |
|---|---|
| Final grades (AA, BA, ... FF) that you get at the end of this semester for this course. | Ordinal ✔ |
| Area code of phone numbers. | Nominal ✔ |
| Number of goals scored by each team in the Turkish Super League | Discrete ✔ |
| Weights of people in our class | Continuous ✔ |

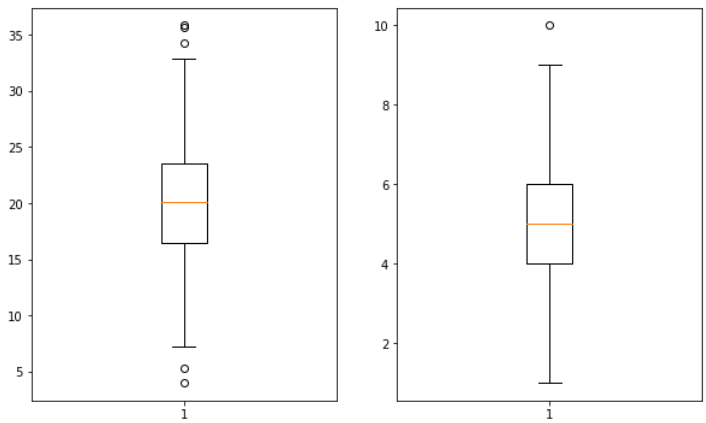
The correct answer is:
Final grades (AA, BA, ... FF) that you get at the end of this semester for this course. → Ordinal,
Area code of phone numbers. → Nominal,
Number of goals scored by each team in the Turkish Super League → Discrete,
Weights of people in our class → Continuous

Question **11**

Correct

1.00 points out of 1.00

Which of the following/s **cannot be said certainly** about the box plots shown below? You can choose more than one.




- [ ] a. Interquartile range is less than 10 in the first (left) boxplot.
- [ ] b. There are outliers in both datasets.
- [x] c. They are both coming from a continuous variable. ✔
- [x] d. Second (right) boxplot have less outliers, so it is more likely coming from a normal distribution. ✔

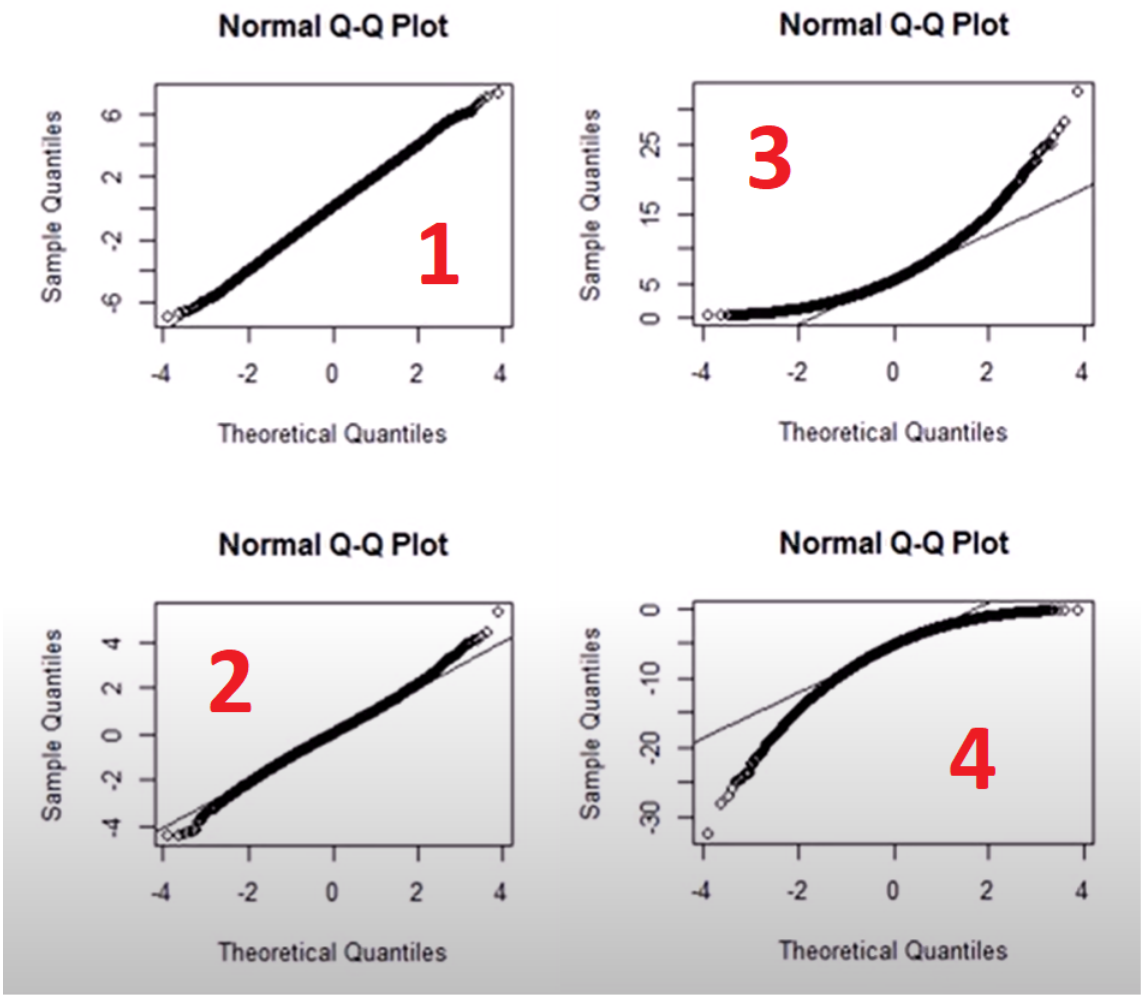
The correct answers are:
They are both coming from a continuous variable.,
Second (right) boxplot have less outliers, so it is more likely coming from a normal distribution.

Question **12**

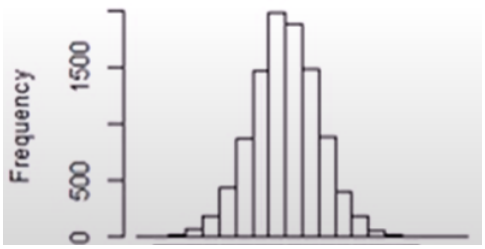
Correct

1.00 points out of 1.00

Some qq plots are given below. Please match them with the distributions below by simply selecting the numbers on the qq-plots in boxes.

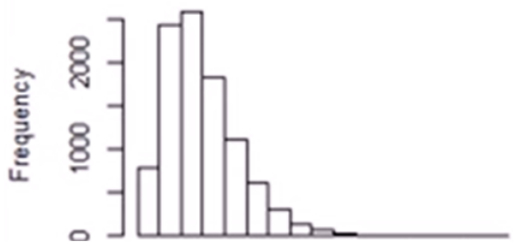



**Symmetric with fat tails**



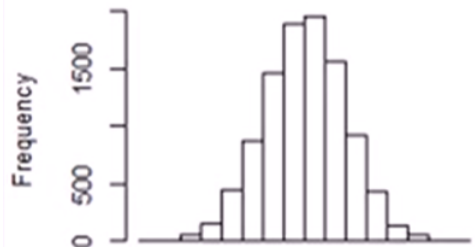

2 ✔

**Positive skew**

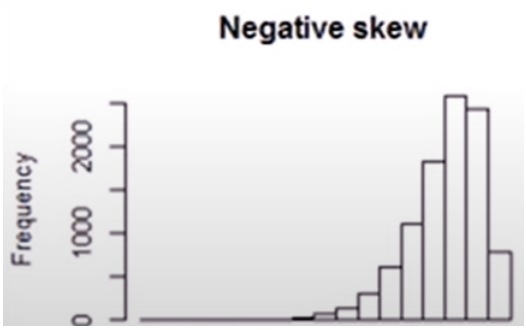



3 ✔

**Symmetric distribution**

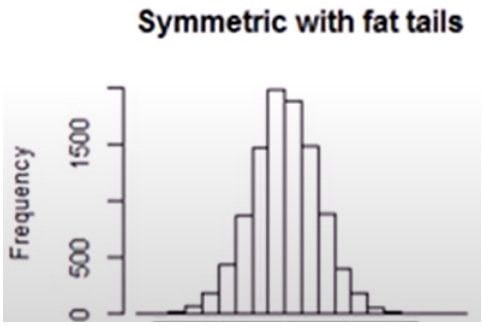



1 ✔

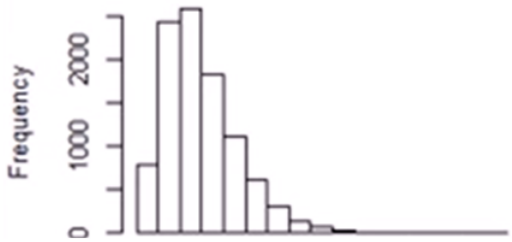



4 ✔

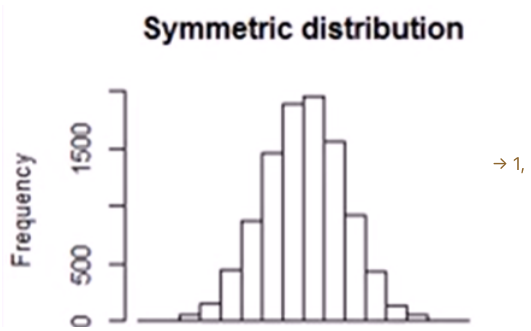
The correct answer is:



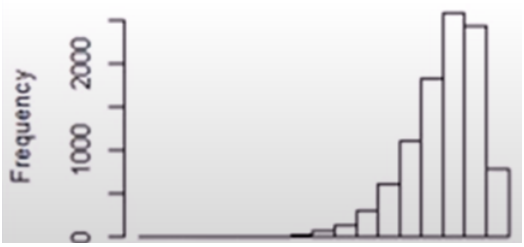

→ 2,




→ 3,




→ 1,




→ 4

◄ Quiz #1

Jump to...

Quiz #3 ►