✓

Congratulations! You passed!
To Pass 80% or higher
Keep Learning
Grade
91.66%

Week 1 Quiz

Latest Submission Grade
91.66%

1.Question 1

At what stage(s) of Data Exploration would you address missing values in a data set?

1 / 1 point

○

Data transformation

◉

Data clean-up

○

Data reduction

✓
Correct

Refer to the following video for a refresher: video 1.

2.Question 2

Which of the following statements regarding data transformation and data reduction is correct?

1 / 1 point

◉

Data transformations work on individual variables, while data reduction works on a set of variables

○

Only data transformation would create dummy variables

○

The goal of data transformations is to create larger datasets while the goal of data reduction is to create smaller datasets

○

Data transformations are out of style; data reduction is the modern man’s tool

✓
Correct

Refer to the following video for a refresher: video 1.

3.Question 3

What does a data value measure after centering and scaling has been applied?

1 / 1 point

○

Accuracy

○

The number of standard deviations between each data point and the median

◉

The number of standard deviations between each data point and the mean

Slope

✓ Correct

Refer to the following video for a refresher: video 1.

4.Question 4

Why would one want to center and scale a set of data?

1 / 1 point

So multiple variables in the data set are on a common scale

To make all data values positive

To remove duplicates

To make data easier to interpret

✓ Correct

Refer to the following video for a refresher: video 1.

5.Question 5

For the following three questions, match the Box-Cox Transformation associated with the given value of lambda.

When Lambda = 0, transformation is

1 / 1 point

Logarithmic

Cubed polynomial

Inverse

Square root

✓ Correct

Refer to the following video for a refresher: video 1.

6.Question 6

When Lambda = 0.5, transformation is

1 / 1 point

Logarithmic

Cubed polynomial

Inverse

Square root

Correct

Refer to the following video for a refresher: video 1.

7.Question 7

When Lambda = - 1, transformation is

1 / 1 point

Logarithmic

Cubed polynomial

Inverse

Square root

Correct

Refer to the following video for a refresher: video 1.

8.Question 8

What is the purpose of applying a Data Reduction?

1 / 1 point

To generate a larger set of variables

To make all variables positively valued

To use a smaller set of variables to capture most of the information in the original variables

Correct

Refer to the following video for a refresher: video 1.

9.Question 9

What must be done to variables of a data set before applying principal component analysis and why?

1 / 1 point

You must scale the variables so that only outliers are considered as principal components

You must scale the variables so that principal components are not dominated by variables of much larger scale

You must make all variables negative to work with values of the same sign

You must take the square root of all data values to reduce the overall magnitudes of the data set

Correct

Refer to the following video for a refresher: video 1.

10.Question 10

Which of the following can be an appropriate way to deal with missing values? (Select all that apply.)

1 / 1 point

Removing the columns or rows with missing values.

Correct

Refer to the following video for a refresher: video 2.

Imputing a value with averages of all other records.

Correct

Refer to the following video for a refresher: video 2.

Imputing a value from “similar” data points.

Correct

Refer to the following video for a refresher: video 2.

Making “missing” its own category.

Correct

Refer to the following video for a refresher: video 2.

11.Question 11

Your organization asks you to analyze a dataset that shows the number of FreeFly ALTA drones sold in 2016. You noticed that only 2 drones were sold the day after Black Friday, while the average number of drones sold in 2016 is around 100 a day. What is the most probable explanation for this small data value?

1 / 1 point

It’s a missing value that someone filled in with a guess

There was a glitch in the system and the data value was corrupted

It’s a censored value that was inputted incorrectly

It's a censored value; drone inventory probably ran out

✓
Correct

Refer to the following video for a refresher: video 2.

12.Question 12

What are the risks of replacing a missing value with a guess? (Select all that apply.)

1 / 1 point

☐ ▢

None, the database is capable of correcting input mistakes

☑ ▢

Introducing biases

✓
Correct

Refer to the following video for a refresher: video 2.

☑ ▢

Distorting the data set

✓
Correct

Refer to the following video for a refresher: video 2.

☐ ▢

Falsifying results

13.Question 13

Why removing all data records with missing values is often not a good way to deal with missing values? (Select all that apply.)

1 / 1 point

☐ ▢

Some modeling tools require a data value for each row/column.

☐ ▢

A dataset is incomplete if there are missing values.

☑ ▢

We may end up with too little data to conduct meaningful analysis.

✓
Correct

Refer to the following video for a refresher: video 2.

☑ ▢

The pattern of missing values can have high predictive power.

✓
Correct

Refer to the following video for a refresher: video 2.

14.Question 14

The table below shows the number of patients visiting a clinic on ten consecutive business days.

Date	Weekday	Patient Count

6/23/14	Monday	16
6/24/14	Tuesday	12
6/25/14	Wednesday	12
6/26/14	Thursday	14
6/27/14	Friday	11
6/30/14	Monday	12
7/1/14	Tuesday	14
7/2/14	Wednesday	11
7/3/14	Thursday	15
7/4/14	Friday	

On July 4, 2014, the clinic is closed to observe the Independence Day. We would like to impute a value for patient count on July 4, 2014. Which of the following are reasonable imputed values? (Select all that apply)

0.75 / 1 point

☑ ☐

0

! This should not be selected

Imputing a value of zero will skew the data.

Refer to the following video for a refresher: video 2.

☑ ☐

13

✓ Correct

Refer to the following video for a refresher: video 2.

☑ ☐

11

✓ Correct

Refer to the following video for a refresher: video 2.

☐ ☐

15.7

15.Question 15

What are the characteristics of an outlier? (Select all that apply.)

0.5 / 1 point

☑ ☐

It is the data point most proximal to the mean

! This should not be selected

Refer to the following video for a refresher: video 3.

☑ ☐

It is the pivot point for the overall pattern that the data follows

! This should not be selected

Refer to the following video for a refresher: video 3.

☑ ☐

It falls far outside the overall data pattern

✓
Correct

Refer to the following video for a refresher: video 3.

☑ ☐

It is above or below 3 standard deviations of the mean

✓
Correct

Refer to the following video for a refresher: video 3.

16.Question 16

A data point is not considered an outlier unless it deviates dramatically on either the x-axis or the y-axis?

1 / 1 point

○ ○

True

◉ ○

False

✓
Correct

Refer to the following video for a refresher: video 3.

17.Question 17

Why do outliers exist? (Select all that apply.)

1 / 1 point

☑ ☐

Data recording errors

✓
Correct

Refer to the following video for a refresher: video 3.

☑ ☐

Legitimate but odd observations

✓
Correct

Refer to the following video for a refresher: video 3.

☐ ☐

Entropy of a system

☐ ☐

Distortion of time

18.Question 18

Which statistical measure is more resistant to outliers?

0 / 1 point

○ ○

Mean

○ ○

Median

Standard deviation

Range

Incorrect

Because standard deviation uses the mean in its calculation, it is not resistant to outliers.

Refer to the following video for a refresher: video 3.

19.Question 19

To say a variable is degenerate means which of the following: (Select all that apply.)

1 / 1 point

The variable is immoral and corrupt

The variable can only take on a single value

Correct

Refer to the following video for a refresher: video 4.

When plotted, the variable is modeled with an exponential decay

The variable is a zero variance variable

Correct

Refer to the following video for a refresher: video 4.

20.Question 20

Which of the following is a remedy to collinearity issues in regression analysis?

1 / 1 point

Adding more dummy variables

Cut the data set in half

Removing zero variance and near zero variance variables

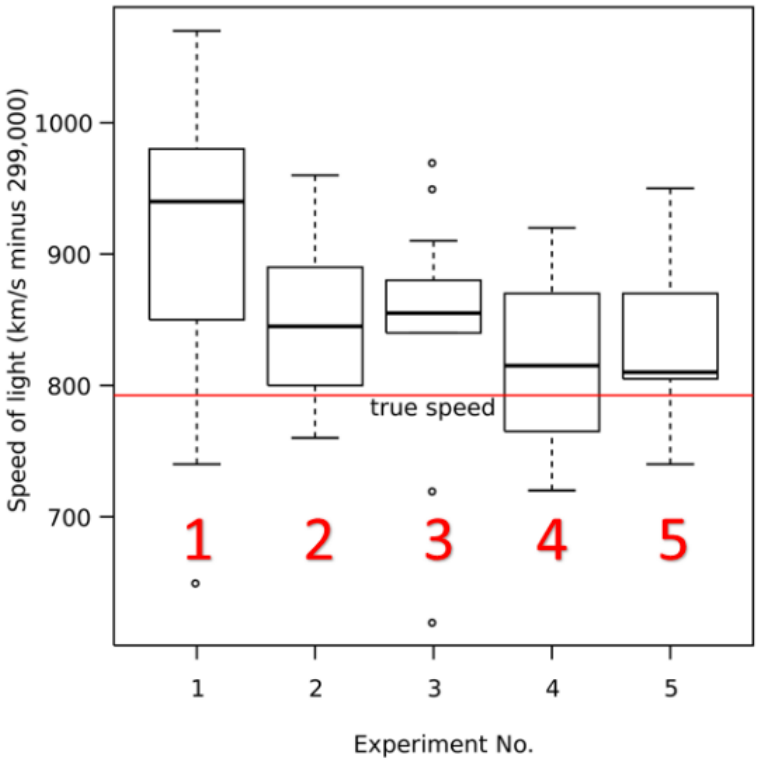
Duplicate the data set

Correct

Refer to the following video for a refresher: video 4.

21.Question 21

Check all valid observations about the set of box plots below: (Select all that apply.)



0.75 / 1 point

☐ ☐ The minimum of Box 4 is less than the minimum of Box 5

☑ ☐ The median of Box 1 is greater than the median of all other boxes

✓ Correct

Refer to the following video for a refresher: video 5.

☑ ☐ Box 3 has a larger median than Box 4

✓ Correct

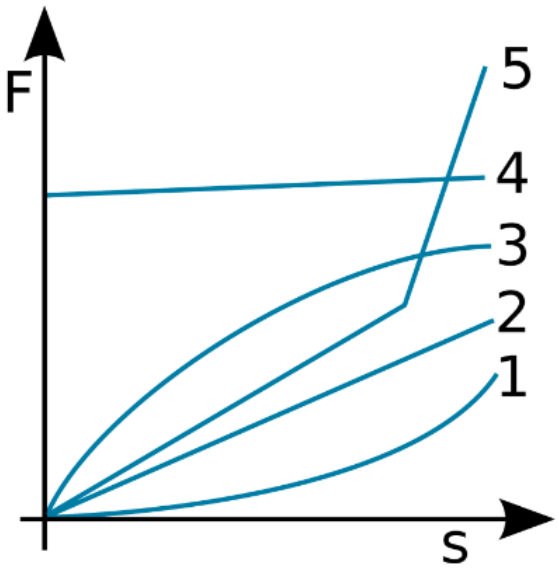
Refer to the following video for a refresher: video 5.

☐ ☐ Box 3 has the largest range of all boxes

You didn't select all the correct answers

22.Question 22

Which function on the graph is both linear and displays a positive relationship?



1 / 1 point

1

2

3

4

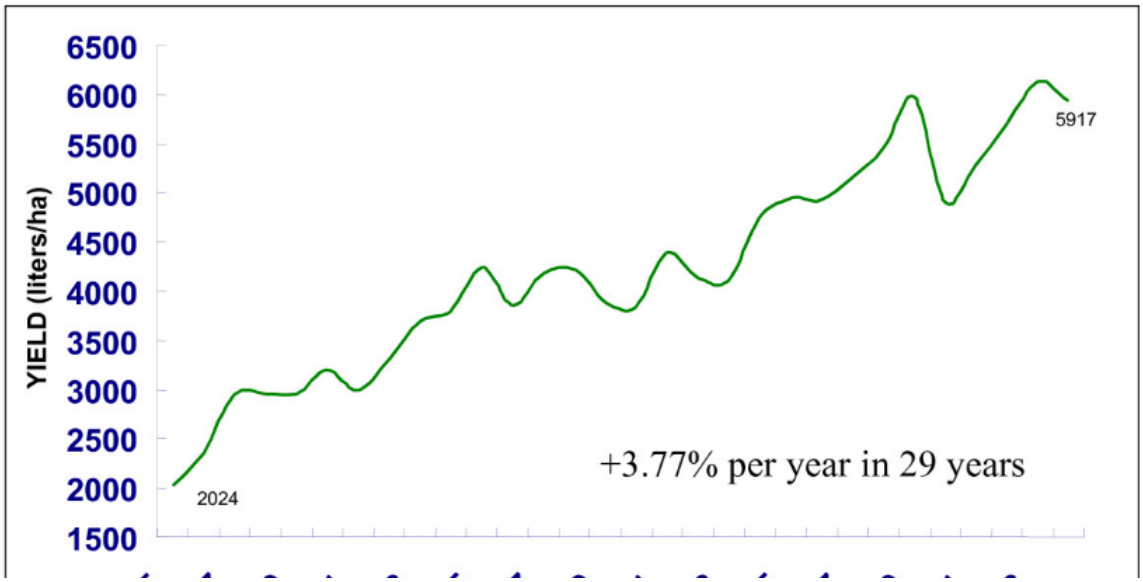
5

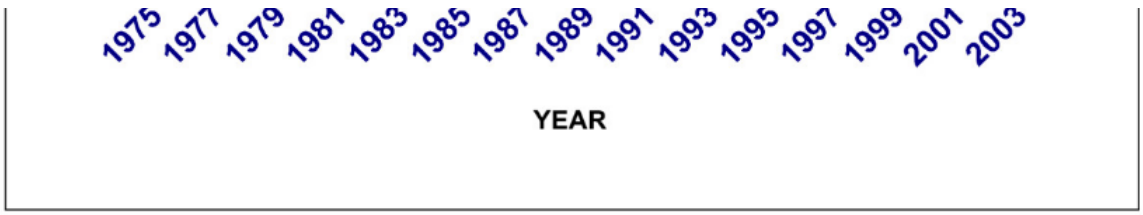
Correct

Refer to the following video for a refresher: video 5.

23.Question 23

How could this time series graph be improved for data visualization purposes?





1 / 1 point

Swap the x and y axis

Adding a line to highlight trend

Convert it to a pie chart

Shade the background with a visually appealing texturized color

Correct

Refer to the following video for a refresher: video 5.

24.Question 24

Given the table below, how many dummy variables should you create based on the number of categories present?

Types of Businesses
Sole Proprietorship
Partnership
Corporation
Limited Liability Company (LLC)

1 / 1 point

1

2

3

4

Correct

In general, a categorical variable with m categories can be represented by m-1 dummy variables.