QUESTION 1

Multiple choice question (1/1 MARKS)

What is the output of the following statement in python?

True or False

- A. True ✓
- B. False

Bravo! Correct answer.

Solution

Correct answer : A

Your answer : A

Explanation

The 'or operation returns True if either condition is True. It returns False when both are False. Therefore, in this case, the output is True.

QUESTION 2

Multiple choice question (1/1 MARKS)

(1,2,3,4,5) is a....

- A. List
- B. Tuple ✓
- C. Set
- D. Array

Excellent! Correct answer.

Solution

Correct answer : B

Your answer : B

Explanation

The values are bounded by a normal bracket this is a tuple.

QUESTION 3

Multiple choice question (1/1 MARKS)

List and Tuple are the same except....

- [x] **A.** Tuple cannot be edited
- [ ] **B.** List cannot be edited
- [ ] **C.** List has fixed length
- [ ] **D.** Tuple has fixed length

Bravo! Correct answer.

Solution

Correct answer : A

Your answer : A

Explanation

Tuple cannot be edited but lists can be.

QUESTION 4

Multiple choice question (1/1 MARKS)

Linspace cannot divide a linear line into a given number of segments.

- A. True
- B. False

Bravo! Correct answer.

Solution

Correct answer : B

Your answer : B

Explanation

Linspace in NumPy divides a linear segment into a given number of segments. For instance np.linspace(1,10,10) returns and array

array([1., 2., 3., 4., 5., 6., 7., 8., 9., 10.])

QUESTION 5

Multiple choice question (0/1 MARKS)

Which of the following movies is the lowest rated movie in the movies dataset?

- A. Proud American ✓
- B. Hopeful Notes
- C. The final wish
- D. Glitter ✗

Incorrect answer.

Solution

Correct answer : A

Your answer : D

Explanation

The lowest rated movies can be found using

```
df[df.Rating == df.Rating.min()]
```

Proud American and Browncoats: Independence War were the lowest rated movies.

QUESTION 6

Multiple choice question (1/1 MARKS)

What was the maximum rating given to any movie?

- A. 10.0
- B. 9.0
- C. 9.9
- D. 9.7 ✓

Well done! Correct answer.

Solution

Correct answer : D

Your answer : D

Explanation

Maximum rating given to any movie can be found using

```
df.Rating.max()
```

The max rating is 9.7

QUESTION 7

Multiple choice question (2/2 MARKS)

What was the highest rating given to any movies in the year 2018?

- A. 9.0
- B. 9.2 ✓
- C. 9.7
- D. 9.1

Perfect! You got this right.

Solution

Correct answer : B

Your answer : B

Explanation

In the year 2018, the highest rating given to any movie was 9.2.

This can be found using the following code

```
df[df.Year == 2018].sort_values('Rating', ascending = False)
```

QUESTION 8

Multiple choice question (2/2 MARKS)

What is the name of the movie director who directed the longest movie in the year 2011?

- A. Ron Maxwell
- B. Matt Lang ✓
- C. Patrick Fogarty
- D. Fred Ashman

Bravo! Correct answer.

Solution

Correct answer : B

Your answer : B

Explanation

The longest movie released in the year 2011 was Starship and the name of the director of the movie is Matt Lang.

This can be found using

```
df[df.Year == 2011].sort_values('Duration', ascending = False)
```

QUESTION 9

Multiple choice question (1/1 MARKS)

Which movie director directed more than 10 movies?

- A. John Lyde
- B. John Swanberg
- C. James Franco
- D. Tyler Perry ✓

Perfect! You got this right.

Solution

Correct answer : D

Your answer : D

Explanation

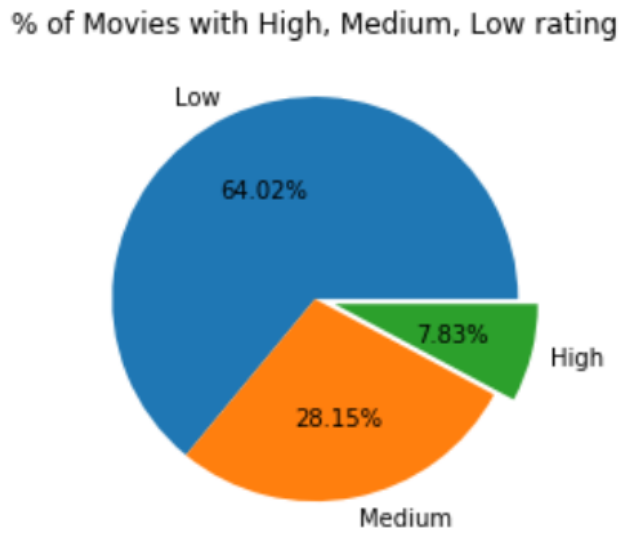
Tyler Perry is the only movie director in the database with more than 10 movies.

This can be found using the following code:

```
df['Director'].value_counts()
```

Multiple choice question (1/1 MARKS)

The following pie chart shows that approximately 28% of movies are highly rated.



A. True

B. False

That's right!

Solution

Correct answer : B

Your answer : B

Explanation

The pie chart shows that 7.83 % of the movies are highly rated.

QUESTION 11

Multiple choice question (1/1 MARKS)

A box plot shows the median value.

- [x] **A.** True
- [ ] **B.** False.

Well done! Correct answer.

Solution

Correct answer : A

Your answer : A

Explanation

A box plot shows min, max, median, 25%, and 75% values and not the mean value.

QUESTION 12

Multiple choice question (1/1 MARKS)

Median can be seen from which of the following univariate charts?

- [x] **A.** Boxplot
- [ ] **B.** Bar chart
- [ ] **C.** Histogram
- [ ] **D.** Density plot

Excellent! Correct answer.

Solution

Correct answer : A

Your answer : A

Explanation

Box plot shows the median values, while histogram and density plots show only the dispersion. The bar chart shows the frequency or a given value per categorical variable.

QUESTION 13

Multiple choice question (1/1 MARKS)

$P(A \cup B) = P(A).P(B)$

- A. True
- B. False ✓

Excellent! Correct answer.

Solution

Correct answer : B

Your answer : B

Explanation

$P(A \cup B) = PA + PB - P(A \cap B)$, if A and B are not mutually exclusive and $P(A \cup B) = P(A) + P(B)$ if A and B are not mutually exclusive.

Two dice are thrown together.

What is the probability of getting a 5 on both dice?

- A. 1/6
- B. 2/6
- C. 1/36 ✓
- D. 2/36

Bravo! Correct answer.

Solution

Correct answer : C

Your answer : C

Explanation

If two dices are thrown together then the sample space is as follows:

(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)
(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)
(3,1), (3,2), (3,3), (3,4), (3,5), (3,6)
(4,1), (4,2), (4,3), (4,4), (4,5), (4,6)
(5,1), (5,2), (5,3), (5,4), **(5,5)**, (5,6)
(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)

We can see that there are 36 possible ways in the sample space and 5 on both dice which is (5,5), can occur in only one possible way.

Therefore, the probability is 1/36

What is the probability of getting either 2 or 4 in at least one of the dice?

- A. 2/36
- B. 20/36 ✓
- C. 12/36
- D. 8/36

Bravo! Correct answer.

Solution

Correct answer : B

Your answer : B

Explanation

(1,1), **(1,2)**, (1,3), **(1,4)**, (1,5), (1,6)

(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)

(3,1), **(3,2)**, (3,3), (**3,4**), (3,5), (3,6)

(4,1), (4,2), (4,3), (4,4), (4,5), (4,6)

(5,1), **(5,2)**, (5,3), (**5,4**), (5,5), (5,6)

(6,1), **(6,2)**, (6,3), (**6,4**), (6,5), (6,6)

We can see from the above sample space that there are 20 possible ways in which either 2 or 4 can show.

Therefore, the probability is 20/36

QUESTION 16

Multiple choice question (1/1 MARKS)

In a normal distribution, what is the probability of the tails beyond 2σ?

- A. 99.73%
- B. 95.45%
- C. 4.55% ✓
- D. 0.27%

That's right!

Solution

Correct answer : C

Your answer : C

Explanation

The probability of the curve within 2σ in a normal distribution is 95.45%. Therefore, the probability outside this region is 4.55%

QUESTION 17

Multiple choice question (2/2 MARKS)

A normally distributed random variable has a mean of 150 and a standard deviation of 21.

What is the probability of getting a value <= 120 from this distribution?

- A. 0.923
- B. 0.745
- C. 0.254
- D. 0.077 ✓

Excellent! Correct answer.

Solution

Correct answer : D

Your answer : D

Explanation

The probability of getting a value <= 120 from this distribution can be found by estimating using the following code:

```
stats.norm.cdf(x = 120,loc = 150, scale = 21)
```

QUESTION 18

Multiple choice question (2/2 MARKS)

A random variable follows t distribution with a mean of 9 and a standard deviation of 3.5.

What is the probability of getting a value >= 10 from this distribution?

- **A.** 0.389 ✓
- **B.** 0.6109
- **C.** 0.613
- **D.** 0.3875

Excellent! Correct answer.

Solution

Correct answer : A

Your answer : A

Explanation

The probability of getting a value >= 10 from this distribution can be found by estimating using the following code:

```
1-stats.t.cdf(x = 10,loc = 9, scale = 3.5, df = 19)
```

QUESTION 19

Multiple choice question (0/2 MARKS)

A random variable follows t distribution with a mean of 9 and a standard deviation of 3.5. Greater than what value is the probability under the curve 0.05?

- A. 3.25
- B. 13.75
- C. 9
- D. 14.75

Incorrect answer.

Solution

Correct answer : D

Your answer : A

Explanation

The value at a certain probability can be computed using the ppf function. The code is given below

```
stats.norm.ppf(q = 0.95, loc = 9, scale = 3.5)
```

QUESTION 20

Multiple choice question (1/1 MARKS)

Which of the following statements is not true about the central limit theorem?

- A. The sampling distribution mean is the same as the population means.
- B. The sampling distribution would be normally distributed.
- C. The sampling distribution is created by sampling the population and calculating the average of each sample set.
- D. The sampling distribution and the population distribution would have the same standard deviation. ✓

Bravo! Correct answer.

Solution

Correct answer : D

Your answer : D

Explanation

The sampling distribution would have a standard deviation of σ/√N Therefore, the standard deviation of the sampling distribution and the population are different.

QUESTION 21

Multiple choice question (1/1 MARKS)

Null hypothesis is not the status quo.

- A. True.
- B. False. ✓

Perfect! You got this right.

Solution

Correct answer : B

Your answer : B

Explanation

In hypothesis testing null hypothesis is the status quo and the alternative hypothesis is the assertion.

QUESTION 22

Multiple choice question (1/1 MARKS)

A certain process had a quality of 99%. The quality manager after sampling a few lots has a suspicion that the quality might have dropped to below 96%

What is the alternative hypothesis in this case?

- A. Ho = 96%
- B. Ho > 99%
- C. Ha > 99%
- D. Ha < 96% ✓

Well done! Correct answer.

Solution

Correct answer : D

Your answer : D

Explanation

The alternative hypothesis is the assertion. Since in this case, the quality manager has a suspicion that the quality might have dropped below 96%, the null and alternative hypotheses are:

Ho >= 96%

Ha < 96%

QUESTION 23

Multiple choice question (0/1 MARKS)

Which of the following statements are true regarding hypothesis testing?

Note: More than one option can be correct.

- [ ] **A.** We cannot reject null hypothesis
- [x] **B.** We can reject the alternative hypothesis
- [x] **C.** We cannot reject alternative hypothesis
- [x] **D.** We can reject the null hypothesis

1/2 answers are incorrect.

Solution

Correct answer : C, D

Your answer : B, D

Explanation

In hypothesis testing, we strive to reject the null hypothesis and when that is done, the alternative hypothesis becomes true. We do not strive to reject the alternative hypothesis in the hypothesis test.

QUESTION 24

Multiple choice question (0/2 MARKS)

100 lots were sampled randomly and found that the average quality was 95% with a standard deviation of 2%.
It was believed that the quality may have dropped to below 90%
What is the Zstat in this case?

- A. +2.5
- B. -2.5
- C. -25 ✓
- D. +25 ✗

Incorrect answer.

Solution

Correct answer : C

Your answer : D

Explanation

Zstat is given by the following formula

$$Zstat = \frac{X - \mu}{\sigma/\sqrt{n}}$$

n this case X = 90, μ = 95, σ = 2 and n = 100

Applying these values in the formula above, the Zstat = -25

QUESTION 25

Multiple choice question (0/1 MARKS)

What is the p-value if the Zstat is calculated to be -25?

- A. 0.006
- B. 1
- C. 0
- D. None of the above

Incorrect answer.

Solution

Correct answer : C

Your answer : D

Explanation

The p-value for a Zstat of -25 would be close to 0.

```
stats.norm.cdf(-25)
```

QUESTION 26

Multiple choice question (1/1 MARKS)

100 samples were collected to test if the mean values have changed from the known mean.

What test should be applied?

- A. T test
- B. Z test ✓
- C. One way ANOVA
- D. Two way ANOVA

Well done! Correct answer.

Solution

Correct answer : B

Your answer : B

Explanation

Since in this case 100 samples are collected, we can use Z distribution. There is no need to do ANOVA since the treatment is not defined as a multi-class categorical variable.

QUESTION 27

Multiple choice question (2/2 MARKS)

A product manager of a popular e-commerce company has a hunch that the cart abandonment would drop if 'Buy Now' flashes once or twice during the online session instead of not flashing at all.

He/She wants to know whether at all it works and whether flashing once is better or twice is better.

What test should be conducted in this case?

- A. Test of means using Z test
- B. Test of means using T test
- C. One-way ANOVA ✓
- D. Two-way ANOVA

Excellent! Correct answer.

Solution

Correct answer : C

Your answer : C

Explanation

In this case, the target variable is the average abandonment rate, and the treatment times 'buy now' is flashed.

Since there are more than 2 groups in this case: 1) no flash, 2) flash once, 3) flash twice, ANOVA must be done as a test of means would be inadequate.

As there is only one treatment there is no need to conduct a two-way ANOVA.

QUESTION 28

Multiple choice question (1/1 MARKS)

Chi-square test ...

- A. Tests if two means are same or not
- B. Tests if two standard deviations are same or not
- C. Tests if two variables are associated or not ✓
- D. Tests if two median values are same or not

That's right!

Solution

Correct answer : C

Your answer : C

Explanation

Chi-square test or test of independence tests if two variables are independent or associated or not.

QUESTION 29

Multiple choice question (2/2 MARKS)

What are the parameters of F-distribution?

- **A.** Degree of freedom of numerator and denominator ✓
- **B.** Degree of freedom
- **C.** Mean and standard deviation
- **D.** None of the above

That's right!

Solution

Correct answer : A

Your answer : A

Explanation

The parameters of the F-distribution are the degree of freedom of the numerator and the denominator, while mean and standard deviation are the parameters for the Z distribution, and the degree of freedom alone of a table is the parameter for chi-square distribution.

QUESTION 30

Multiple choice question (2/2 MARKS)

The pvalue for a given Fstat is 0.02 in a one way ANOVA. What does this mean?

- A. The group means are all the same
- B. The variables are independent
- C. The group means are different ✓
- D. The group means are lesser than the given value

Excellent! Correct answer.

Solution

Correct answer : C

Your answer : C

Explanation

Since the pvalue is 0.02 in a one way ANOVA, one can reject the null hypothesis, which is that the group means are all the same.

Therefore, the group means are different.

QUESTION 31

Multiple choice question (1/1 MARKS)

What is the null hypothesis in a Chi-square test?

- **A.** The variables are not associated ✓
- **B.** The group means are same
- **C.** The group means are not the same
- **D.** The groups are associated

Well done! Correct answer.

Solution

Correct answer : A

Your answer : A

Explanation

The null hypothesis in a chi-square test is that the variables are not associated.

QUESTION 32

Multiple choice question (0/1 MARKS)

How is the degree of freedom in a Chi-square test calculated?

- A. (1-rows)*(1-columns) ✓
- B. N-1
- C. N-k-1
- D. None of the above ✗

Incorrect answer.

Solution

Correct answer : A

Your answer : D

Explanation

Degree of freedom in a Chi-square test is calculated as:

(1-rows) times (1-columns)

QUESTION 33

Multiple choice question (1/1 MARKS)

What is true about Tukey's honest significance difference test?

- A. It is same as ANOVA
- B. It is same as T-test
- C. It is applicable only for two-way ANOVA
- D. It compares the pairwise groups in ANOVA

Well done! Correct answer.

Solution

Correct answer : D

Your answer : D

Explanation

Tukey's honest significance difference test compares the pairwise group means of all groups in ANOVA. Even though it is similar to t-test the computation is different. And it can be used in both one-way as well as two-way ANOVA.

QUESTION 34

Multiple choice question (3/3 MARKS)

A contingency table with 4 rows and 5 columns was constructed. The chi-square statistics were calculated to be 20. Given a significance of 0.05, are the given two variables independent of each other?

- A. Independent
- B. Not independent ✓

Well done! Correct answer.

Solution

Correct answer : B

Your answer : B

Explanation

In this case the degree of freedom is (4-1)*(5-1) = 3*4 = 12

Given the chi-square statistics is 20, the p-value can be calculated as 0.067 using the following code

```
1-stats.chi2.cdf(20,df=12)
```

We cannot reject the null hypothesis with a p-value of 0.067 against the significance value of 0.05.

Therefore, the conclusion is that variables are not independent.

QUESTION 35

Multiple choice question (1/1 MARKS)

What is the main objective of regression analysis?

- A. To test whether the means are different across groups
- B. To study which variables are not associated
- C. To predict a continuous variable using existing relationships with other variables ✓
- D. To study which variables are associated

Bravo! Correct answer.

Solution

Correct answer : C

Your answer : C

Explanation

The objective of linear regression is to predict a continuous variable using existing relationships with other variables

QUESTION 36

Multiple choice question (0/1 MARKS)

The regression analysis cannot predict a categorical variable.

- A. True ✓
- B. False ✗

Incorrect answer.

Solution

Correct answer : A

Your answer : B

Explanation

Regression is used to predict the continuous variable and not categorical variables.

QUESTION 37

Multiple choice question (1/1 MARKS)

What measure shows how good a regression model is?

- A. β
- B. F-stat
- C. pvalue
- D. R-square ✓

That's right!

Solution

Correct answer : D

Your answer : D

Explanation

R-square of goodness of fit is a measure of how good a model is.

QUESTION 38

Multiple choice question (1/1 MARKS)

In a simple linear regression model, the square of the correlation coefficient is exactly the same as the goodness of fit.

- A. True ✓
- B. False

Excellent! Correct answer.

Solution

Correct answer : A

Your answer : A

Explanation

In a simple linear regression model, the square of the correlation coefficient is exactly the same as the goodness of fit.

QUESTION 39-41

Reading Comprehension (4/4 MARKS)

A multiple linear regression model is given as:

Calories = 6.12 + 11.3 * TransFat + 4 * Sugars + 6.1 * Protein

QUESTION 39 (1/1 MARK)

What is the βo in this model?

- A. 11.3
- B. 4
- C. 6.1
- D. 6.12 ✓

Solution

Correct answer : D

Your answer : D

Explanation

βo is the intercept, which is 6.12 in this case.

QUESTION 40 (1/1 MARK)

By how much would Calories increase if transfat increased by half a gram?

- A. 2
- B. 4
- C. 5.65 ✓
- D. 11.3

Solution

Correct answer : C

Your answer : C

Explanation

Calories increases by 5.65 if TransFat increases by 0.5 as the coefficient for transfat is 11.3

QUESTION 41 (2/2 MARK)

If transfat is 0.1, sugars is 9 and protein is 6, then what are the total calories?

- A. 80 ✓
- B. 260
- C. 90
- D. 70

Solution

Correct answer : A

Your answer : A

Explanation

Applying transfat as 0.1, sugars as 9 and protein as 6 in the following equation;

Calories = 6.12 + 11.3 * 0.1 + 4 * 9 + 6.1 * 6 = 79.85

Correct Answer.

QUESTION 42

Multiple choice question (1/1 MARKS)

Which of the following variables are not significant in a linear model predicting the house price?

- [x] **A.** chas
- [ ] **B.** nox
- [ ] **C.** rm
- [ ] **D.** dis

Perfect! You got this right.

Solution

Correct answer : A

Your answer : A

Explanation

The coefficient for the variable Chas has a p-value of 0.34, which means that the variable is not significant.

QUESTION 43

Multiple choice question (2/2 MARKS)

The nitrogen oxide concentration has no impact on the house price.

- A. True
- B. False ✓

Excellent! Correct answer.

Solution

Correct answer : B

Your answer : B

Explanation

From the MLR model built, we can see that the coefficient of nox is -15.5, this means that the price decreases by $15k and the pvalue is almost 0. This means that the nitrogen oxide concentration has a significant as well as a sizeable impact on the house prices.

QUESTION 44

Multiple choice question (2/2 MARKS)

Which of the following variables are least impactful in predicting house prices?

- A. nox ✓
- B. zad
- C. chas
- D. rm

Perfect! You got this right.

Solution

Correct answer : A

Your answer : A

Explanation

The parameters of the regression model using scaled data are given below.

```
Intercept    0.538025
rm           0.419029
rad          0.137338
zn           0.089308
nox         -0.172322
ptratio     -0.185143
tax         -0.186540
crim        -0.239221
dis         -0.276054
lstat       -0.334317
dtype: float64
```

We can see that nox is the second least impactful variable.