X


(https://swayam.gov.in)  (https://swayam.gov.in/nc_details/NPTEL)

sai.doc45@gmail.com

Week 1 : Assignment1

The due date for submitting this assignment has passed.

Due on 2022-02-09, 23:59 IST.

Assignment submitted on 2022-02-06, 17:16 IST

1) State True or false: ***1 point***

Statement: Data can be generated by machines but not by humans.

- ○ True
- ◉ False

Yes, the answer is correct.
Score: 1

Accepted Answers:
False

2) Which one of the following is not a classification of Data Analytics? ***1 point***

- ○ Diagnostic analytics
- ◉ Deceptive analytics
- ○ Predictive analytics
- ○ Prescriptive analytics

Yes, the answer is correct.
Score: 1

Accepted Answers:
Deceptive analytics

3) State True or false: ***1 point***

Statement: Nominal scale is the lowest level of measurement and ratio scale is the highest level of measurement.

- ◉ True
- ○ False

Yes, the answer is correct.
Score: 1
Accepted Answers:
True

4) Consider the following statements- Statement A : With iloc, we can pass in the negative value. Statement B : With loc, we can pass in the negative value. ***1 point***

- A and B are correct
- Both are false
- A is correct B is false
- B is correct A is false

Yes, the answer is correct.
Score: 1
Accepted Answers:
A is correct B is false

5) For getting 4rd, 5th and 7th row of a datafile “df”in Python programming, we can write: ***1 point***

- df.loc[[3,4,6]]
- df.loc[[4,5,7]]
- df.iloc[3,4,6]
- None of these

No, the answer is incorrect.
Score: 0
Accepted Answers:
df.loc[[3,4,6]]

6) Which of the following is not a measure of dispersion? ***1 point***

- Skewness
- Kurtosis
- Range
- percentile

Yes, the answer is correct.
Score: 1
Accepted Answers:
percentile

7) State the following true or false? Statement: Bimodal Data sets contains two modes. ***1 point***

- True
- False

Yes, the answer is correct.
Score: 1
Accepted Answers:
True

8) Bar Charts are used for : ***1 point***

- Continuous data
- Categorical data
- Both of these

problems with us ()

- None of these

Yes, the answer is correct.
Score: 1
Accepted Answers:
Categorical data

9) Median is not applicable to: ***1 point***

- Ordinal data
- Interval data
- Nominal data
- None of these

Yes, the answer is correct.
Score: 1
Accepted Answers:
Nominal data

10) State true or false: Statement: Arithmetic mean is not applicable for ordinal or nominal data ***1 point***

- True
- False

Yes, the answer is correct.
Score: 1
Accepted Answers:
True

X

Week2 : Assignment 2

The due date for submitting this assignment has passed.

Due on 2022-02-09, 23:59 IST.

Assignment submitted on 2022-02-06, 15:48 IST

1) A college plans to interview 8 students for possible offers of graduate assistant-ships. The college has three assistant-ships available. How many groups of three can the college select? ***1 point***

- 126
- 56
- 136
- 130

Yes, the answer is correct.
Score: 1
Accepted Answers:
56

2) Which of the following is each individual outcome of an experiment called? ***1 point***

- the sample space
- a sample point
- an experiment
- an individual

Yes, the answer is correct.
Score: 1
Accepted Answers:
a sample point

3) Two events having nonzero probabilities ***1 point***

- can be both mutually exclusive and independent

- cannot be both mutually exclusive and independent
- are always mutually exclusive
- are always independent

Yes, the answer is correct.
Score: 1
Accepted Answers:
cannot be both mutually exclusive and independent

4) Ten individuals are candidates for positions of president, vice president of an organization. How many possibilities of selections exist? ***1 point***

- 90
- 100
- 120
- 130

Yes, the answer is correct.
Score: 1
Accepted Answers:
90

5) A standard normal distribution has: ***1 point***

- Mean 1 and standard distribution 0
- Mean 0.5 and standard distribution 0.5
- Mean 0 and standard distribution 1
- Mean 1 and standard distribution 1

Yes, the answer is correct.
Score: 1
Accepted Answers:
Mean 0 and standard distribution 1

6) The weight of cows in a farm is normally distributed with a mean of 200 pounds and a standard deviation of 25 pounds. ***1 point***
The probability of a cow weighing more than 241.25 pounds is:

- 0.4505
- 0.0495
- 0.9505
- 0.9010

Yes, the answer is correct.
Score: 1
Accepted Answers:
0.0495

7) The weight of cows in a farm is normally distributed with a mean of 200 pounds and a standard deviation of 25 pounds. ***1 point***
what is the probability of a cow weighing less than 250 pounds?

- 0.4772
- 0.9772
- 0.0528

- 0.5000

Yes, the answer is correct.
Score: 1

Accepted Answers:
0.9772

8) State True or False: ***1 point***

Statement: Binomial and normal Distributions are discrete probability distributions

- True
- False

Yes, the answer is correct.
Score: 1

Accepted Answers:
False

9) For a binomial experiment with p = 0.5 and a sample size of 100. The expected value of this distribution is? ***1 point***

- 0.50
- 0.30
- 100
- 50

Yes, the answer is correct.
Score: 1

Accepted Answers:
50

10) State true or false ***1 point***

Statement: All mutually exclusive events are independent events

- True
- False

Yes, the answer is correct.
Score: 1

Accepted Answers:
False

X


(https://swayam.gov.in)  (https://swayam.gov.in/nc_details/NPTEL)

sai.doc45@gmail.com

Week3 - Assignment 3

The due date for submitting this assignment has passed.

Due on 2022-02-16, 23:59 IST.

Assignment submitted on 2022-02-16, 20:03 IST

1) Sate True or False: ***1 point***

Statement: The specific value of a random variable is called estimator

- (•) True
- ( ) False

No, the answer is incorrect.
Score: 0

Accepted Answers:
False

2) If the true proportion of customers who are below 20 years is P=0.35, what is the probability that a sample size 100 yields a sample proportion between 0.3 to 0.4 ***1 point***

- ( ) 0.961
- ( ) 0.827
- (•) 0.706
- ( ) 0.53

Yes, the answer is correct.
Score: 1

Accepted Answers:
0.706

3) Stratified random sampling is a method of selecting a sample in which ***1 point***

- ( ) the sample is first divided into strata, and then random samples are taken from each stratum
- ( ) various strata are selected from the sample

- the population is first divided into strata, and then random samples are drawn from each stratum
- None of these alternatives is correct

Yes, the answer is correct.
Score: 1
Accepted Answers:
the population is first divided into strata, and then random samples are drawn from each stratum

4) Sate True or False: ***1 point***

Statement: A population is a set of all items or individual of interest

- True
- False

Yes, the answer is correct.
Score: 1
Accepted Answers:
True

5) A question paper contains 90 multiple choice questions. There are 4 alternative answers (A, B, C or D) out of which only one is correct. Mr X answers these questions randomly (i.e. without preparation). What is the probability that X gets a score of at least 10 marks? ***1 point***

- 0.9997
- 0.7894
- 0
- 0.001

No, the answer is incorrect.
Score: 0
Accepted Answers:
0.9997

6) On an average 5 % items supplied by manufacturer X are defectives. If a batch of 10 items is inspected: what is the probability that 2 items are defective ***1 point***

- 0.065
- 0.075
- 0.085
- 0.095

Yes, the answer is correct.
Score: 1
Accepted Answers:
0.075

7) A car distributor in city Y experiences on an average 2.5 car sales per day. Find the probability that on a randomly selected day, they will sell 5 car: ***1 point***

- 0.0668
- 0.544
- 0.082
- 0.205

Live sessions - Solve sample problems with us ()

Yes, the answer is correct.
Score: 1
Accepted Answers:
0.0668

8) A car distributor in city Y experiences on an average 2.5 car sales per day. Find the probability that on a randomly selected day, they will sell no car: ***1 point***

- 0.0668
- 0.544
- 0.082
- 0.205

Yes, the answer is correct.
Score: 1
Accepted Answers:
0.082

9) A random sample of 100 people shows that 25 of them are females and rest are males. Form a 95% confidence interval for the true proportion of females. The lower limit of this interval will be: ***1 point***

- 0.150
- 0.145
- 0.165
- 0.175

Yes, the answer is correct.
Score: 1
Accepted Answers:
0.165

10) A random sample of 100 people shows that 25 of them are females and rest are males. Form a 95% confidence interval for the true proportion of females. The upper limit of this interval will be: ***1 point***

- 0.150
- 0.165
- 0.465
- 0.335

Yes, the answer is correct.
Score: 1
Accepted Answers:
0.335

X

Week4 : Assignment 4

The due date for submitting this assignment has passed.

Due on 2022-02-23, 23:59 IST.

Assignment submitted on 2022-02-23, 09:52 IST

1) If we have a sample size of 20 and population standard deviation is known, we will use: ***1 point***

- t- test for hypothesis testing
- z-test for hypothesis testing
- both t and z test
- F-test

Yes, the answer is correct.
Score: 1

Accepted Answers:
z-test for hypothesis testing

2) Null hypothesis, Ho: mu1- mu2 = 0 is a: ***1 point***

- Upper tail test
- Lower tail test
- Two tail test
- F-test

Yes, the answer is correct.
Score: 1

Accepted Answers:
Two tail test

3) The quality-control manager at a Li-BATTERY factory needs to determine whether the mean life of a large shipment of Li-Battery is equal to the specified value of 375 hours. The ***1 point***

process standard deviation is known to be 100 hours. A random sample of 64 batteries indicates a sample mean life of 350 hours.
State the null and alternative hypotheses

- (●) Mu = 375
- ( ) Mu ≤ 375
- ( ) Mu = 350
- ( ) Mu ≥ 350

Yes, the answer is correct.
Score: 1
Accepted Answers:
Mu = 375

4) The quality-control manager at a Li-BATTERY factory needs to determine whether the mean life of a large shipment of Li-Battery is equal to the specified value of 375 hours. The process standard deviation is known to be 100 hours. A random sample of 64 batteries indicates a sample mean life of 350 hours. ***1 point***
At the alpha = 0.05 level of significance is there any evidence that the mean life is different from 375 hours?

- ( ) Yes, there is
- (●) No, there is not
- ( ) None of these

No, the answer is incorrect.
Score: 0
Accepted Answers:
Yes, there is

5) For one-tailed test, the test statistic z is determined to be zero. The p-value for this test is: ***1 point***

- ( ) Zero
- ( ) -0.5
- (●) +0.5
- ( ) 1.0

Yes, the answer is correct.
Score: 1
Accepted Answers:
+0.5

6) The error of rejecting a true null hypothesis is: ***1 point***

- (●) a Type I error
- ( ) a Type II error
- ( ) is the same as Beta
- ( ) committed when not enough information is available

Yes, the answer is correct.
Score: 1
Accepted Answers:
a Type I error

7) The mean cost of a hotel room in a city is said to be $168 per night. A random sample of 25 hotels resulted in X-bar = $172.50 and sample standard deviation s = 15.40. Calculate the t statistic. **1 point**

- 2
- -2
- 1.46
- -1.46

Yes, the answer is correct.
Score: 1

Accepted Answers:
1.46

8) In hypothesis testing if the null hypothesis is rejected, **1 point**

- no conclusions can be drawn from the test
- the alternative hypothesis is true
- the data must have been accumulated incorrectly
- the sample size has been too small

Yes, the answer is correct.
Score: 1

Accepted Answers:
the alternative hypothesis is true

9) In the hypothesis testing procedure, alpha is **1 point**

- 1 - the level of significance
- the critical value
- the confidence level
- the level of significance

Yes, the answer is correct.
Score: 1

Accepted Answers:
the level of significance

10) If a hypothesis is rejected at the 5% level of significance, it **1 point**

- will always be rejected at the 1% level
- will always be accepted at the 1% level
- will never be tested at the 1% level
- May be rejected or not rejected at the 1% level

No, the answer is incorrect.
Score: 0

Accepted Answers:
May be rejected or not rejected at the 1% level

X


(https://swayam.gov.in)  (https://swayam.gov.in/nc_details/NPTEL)

sai.doc45@gmail.com

Week5 : Assignment 5

The due date for submitting this assignment has passed.

Due on 2022-03-02, 23:59 IST.

Assignment submitted on 2022-03-02, 10:07 IST

1) In the analysis of variance procedure (ANOVA) the term "factor" refers to: ***1 point***

- the dependent variable
- the independent variable
- different levels of a treatment
- the critical value of F

Yes, the answer is correct.
Score: 1
Accepted Answers:
the independent variable

2) In a problem of ANOVA, involving 3 treatments and 10 observations per treatment, SSE = 500. The MSE for this situation is: ***1 point***

- 130.2
- 48.8
- 18.52
- 30.0

Yes, the answer is correct.
Score: 1
Accepted Answers:
18.52

3) The 'F' ratio in a completely randomized ANOVA is the ratio of ***1 point***

- MST/MSE
- MSTR/MSE

- MSE/MSTR
- MSE/MST

Yes, the answer is correct.
Score: 1
Accepted Answers:
MSTR/MSE

4) An ANOVA procedure is applied to data obtained from 7 samples where each sample contains 10 observations. The degrees of freedom for the critical value of F are: ***1 point***

- 7 numerator and 20 denominator degrees of freedom
- 5 numerator and 20 denominator degrees of freedom
- (6 numerator and 63 denominator degrees of freedom)
- 7 numerator and 63 denominator degrees of freedom

Yes, the answer is correct.
Score: 1
Accepted Answers:
6 numerator and 63 denominator degrees of freedom

5) In an ANOVA problem if SST = 200 and SSTR = 80, then SSE is ***1 point***

- 280
- (120)
- 80
- 220

Yes, the answer is correct.
Score: 1
Accepted Answers:
120

6) The critical F value with 8 numerator and 29 denominator degrees of freedom at alpha = 0.01 is ***1 point***

- 2.18
- (3.20)
- 3.53
- 3.94

Yes, the answer is correct.
Score: 1
Accepted Answers:
3.20

7) Two Independent simple random samples are taken to test the difference between the means of two populations. The standard deviations are not known, but are assumed to be equal. The sample sizes are n1 = 15 and n2 = 35. The correct distribution to use is the: ***1 point***

- t distribution with 51 degrees of freedom
- z distribution with 50 degrees of freedom
- z distribution with 49 degrees of freedom
- (t distribution with 48 degrees of freedom)

Yes, the answer is correct.
Score: 1

Accepted Answers:
t distribution with 48 degrees of freedom

8) Stare true or false: ***1 point***

Statement: The sampling distribution of two populations P bar1 -P bar2 is approximated by a normal distribution

- (●) True
- ( ) False

Yes, the answer is correct.
Score: 1

Accepted Answers:
True

9) Mean marks obtained by male and female students of school ABCD in first unit test ***1 point*** are shown as below.

	Male	Female
Sample Size	64	36
Sample Mean Marks	44	41
Population Variance	128	72

The standard error for the difference between the two means is

- ( ) 4.0
- ( ) 7.46
- ( ) 4.24
- (●) 2.0

Yes, the answer is correct.
Score: 1

Accepted Answers:
2.0

10) Mean marks obtained by male and female students of school ABCD in first unit test ***1 point*** are shown as below.

	Male	Female
Sample Size	64	36
Sample Mean Marks	44	41
Population Variance	128	72

If you are interested in testing whether or not the average marks of males is significantly greater than that of females, the test statistic is:

- ( ) 2.0
- (●) 1.5
- ( ) 1.96
- ( ) 1.645

Yes, the answer is correct.
Score: 1

Accepted Answers:
1.5

X


(https://swayam.gov.in)  (https://swayam.gov.in/nc_details/NPTEL)

sai.doc45@gmail.com

NPTEL (https://swayam.gov.in/explorer?ncCode=NPTEL) » Data Analytics with Python (course)

Week6 : Assignment 6

The due date for submitting this assignment has passed.

Due on 2022-03-09, 23:59 IST.

Assignment submitted on 2022-03-09, 19:28 IST

1) Sate True or False: ***1 point***

Statement: In regression analysis the error term is normally distributed

- (•) True
- ( ) False

Yes, the answer is correct.
Score: 1

Accepted Answers:
True

2) In model developed from sample data having the form of "yhat = b0 +b1 x" is known as ***1 point***

- ( ) regression equation
- ( ) correlation equation
- (•) estimated regression equation
- ( ) regression model

Yes, the answer is correct.
Score: 1

Accepted Answers:
estimated regression equation

3) State True or False: Statement: A completely randomized design (CRD) is useful when the experimental units are heterogeneous ***1 point***

- ( ) True
- (•) False

Yes, the answer is correct.
Score: 1

Accepted Answers:
False

4) In a regression and correlation analysis if r2 = 1, then ***1 point***

- SSE = SST
- SSE = 1
- SSR = SSE
- SSR = SST

Yes, the answer is correct.
Score: 1

Accepted Answers:
SSR = SST

5) In a completely randomized design, a random sample of Salesmen would be assigned to each shop alternatively. However, salesmen are believed to differ substantially in their ability to handle number of customers. What is high surge of customers to one salesman might be only moderate or even low surge to another. A study measuring the efficiency of the salesmen resulted in proposals for modification and redesign of the salesmen's work schedule. After consideration of several schedules for the work, three specific alternatives are selected as having the best potential for increasing the efficiency of the salesmen. Check to what extent does the three alternatives differ in terms of their effect on the efficiency of the salesmen? ***1 point***

Salesman	Schedule1	Schedule2	Schedule3
1	75	76	78
2	74	74	74
3	70	71	75
4	73	72	77
5	76	73	76
6	73	73	73

After performing one way ANOVA on the above problem we will:

- Accept the null hypothesis
- Reject the null hypothesis
- Can't state any conclusion
- None of these

Yes, the answer is correct.
Score: 1

Accepted Answers:
Accept the null hypothesis

6) In a completely randomized design, a random sample of Salesmen would be assigned to each shop alternatively. However, salesmen are believed to differ substantially in their ability to handle number of customers. What is high surge of customers to one salesman might be only moderate or even low surge to another. A study measuring the efficiency of the salesmen resulted in proposals for modification and redesign of the salesmen's work schedule. After consideration of several schedules for the work, three specific alternatives are selected as having the best potential for increasing the efficiency of the salesmen. Check to what extent does the three alternatives differ in terms of their effect on the efficiency of the salesmen? ***1 point***

Salesman	Schedule1	Schedule2	Schedule3

1	75	76	78
2	74	74	74
3	70	71	75
4	73	72	77
5	76	73	76
6	73	73	73

After performing one way RBD on this problem we will

- Accept the null hypothesis
- Reject the null hypothesis
- Can't state any conclusion
- None of these

Yes, the answer is correct.
Score: 1
Accepted Answers:
Reject the null hypothesis

7) In a completely randomized design, a random sample of Salesmen would be assigned to each shop alternatively. However, salesmen are believed to differ substantially in their ability to handle number of customers. What is high surge of customers to one salesman might be only moderate or even low surge to another. A study measuring the efficiency of the salesmen resulted in proposals for modification and redesign of the salesmen's work schedule. After consideration of several schedules for the work, three specific alternatives are selected as having the best potential for increasing the efficiency of the salesmen. Check to what extent does the three alternatives differ in terms of their effect on the efficiency of the salesmen? ***1 point***

Salesman	Schedule1	Schedule2	Schedule3
1	75	76	78
2	74	74	74
3	70	71	75
4	73	72	77
5	76	73	76
6	73	73	73

The value of MSE when this problem is solved by ANOVA is:

- 1.955
- 9.555
- 6.855
- 3.588

Yes, the answer is correct.
Score: 1
Accepted Answers:
3.588

8) In a completely randomized design, a random sample of Salesmen would be assigned to each shop alternatively. However, salesmen are believed to differ substantially in their ability to handle number of customers. What is high surge of customers to one salesman might be only moderate or even low surge to another. A study measuring the efficiency of the ***1 point***

salesmen resulted in proposals for modification and redesign of the salesmen's work schedule. After consideration of several schedules for the work, three specific alternatives are selected as having the best potential for increasing the efficiency of the salesmen. Check to what extent does the three alternatives differ in terms of their effect on the efficiency of the salesmen?

Salesman	Schedule1	Schedule2	Schedule3
1	75	76	78
2	74	74	74
3	70	71	75
4	73	72	77
5	76	73	76
6	73	73	73

The value of MSE when this problem is solved by RBD is:

- (●) 1.955
- ( ) 9.555
- ( ) 6.855
- ( ) 3.588

Yes, the answer is correct.
Score: 1

Accepted Answers:
1.955

9) State True of false **1 point**

Statement: The variance of error, is same for all values of the independent variable

- (●) True
- ( ) False

Yes, the answer is correct.
Score: 1

Accepted Answers:
True

10) SSE can never be **1 point**

- (●) larger than SST
- ( ) smaller than SST
- ( ) equal to 1
- ( ) equal to zero

Yes, the answer is correct.
Score: 1

Accepted Answers:
larger than SST

X

Week7 : Assignment 7

The due date for submitting this assignment has passed.

Due on 2022-03-16, 23:59 IST.

Assignment submitted on 2022-03-15, 13:57 IST

1) Which of the following is not an assumption for simple linear regression? ***1 point***

- (•) Normally distributed variables
- ( ) Multicollinearity
- ( ) Linear relationship
- ( ) Normally distributed residuals

No, the answer is incorrect.
Score: 0

Accepted Answers:
Multicollinearity

2) Which of the following is called Standard Error? ***1 point***

- ( ) T-statistic squared
- ( ) Square root of SSE
- ( ) Square root of SST
- (•) Square root of MSE

Yes, the answer is correct.
Score: 1

Accepted Answers:
Square root of MSE

3) Which of the following is true about multiple regression model? ***1 point***

- ( ) It has only one independent variable
- ( ) It has more than one dependent variable
- (•) It has more than one independent variable

It has at least 2 dependent variables

Yes, the answer is correct.
Score: 1

Accepted Answers:
It has more than one independent variable

4) In a multiple regression model, the error term ε is assumed to **1 point**

- Have a mean of 1
- Have a variance of 0
- Have a standard deviation of 1
- Be normally distributed

Yes, the answer is correct.
Score: 1

Accepted Answers:
Be normally distributed

5) For a multiple regression model with 2 independent variables, R.sq = 0.904 and adjusted R. sq = 0.88, determine the number of observations (n) **1 point**

- 6
- 7
- 9
- 10

Yes, the answer is correct.
Score: 1

Accepted Answers:
10

6) If the R.sq value is small for a model with a large number of independent variables, the adjusted coefficient of determination _____ **1 point**

- Can be positive
- Can be negative
- is 0
- Can't say

Yes, the answer is correct.
Score: 1

Accepted Answers:
Can be negative

7) Which one of the statements is true regarding residuals in regression analysis? **1 point**

- Mean of residuals is always 0
- Mean of residuals is always < 0
- Mean of residuals is always > 0
- There is no such rule for residuals

Yes, the answer is correct.
Score: 1

Books ()

Live sessions - Solve sample problems with us ()

Accepted Answers:
Mean of residuals is always 0

8) In a simple linear regression model (one independent variable), if we change the input variable by 1 unit, how much will the output variable change? ***1 point***

- By 1
- No change
- By its slope
- None of these

Yes, the answer is correct.
Score: 1
Accepted Answers:
By its slope

9) To check whether a significant relationship exists between the dependent and set of all independent variables, _____ is used. It is the test for overall significance. ***1 point***

- F-test
- R.sq test
- a correlation test
- t-test

Yes, the answer is correct.
Score: 1
Accepted Answers:
F-test

10) Which of the following evaluation metrics is used to evaluate a regression model? ***1 point***

- AUC-ROC
- Accuracy
- Logloss
- Mean-Squared-Error

Yes, the answer is correct.
Score: 1
Accepted Answers:
Mean-Squared-Error

X


(https://swayam.gov.in)  (https://swayam.gov.in/nc_details/NPTEL)

sai.doc45@gmail.com

Week8 : Assignment 8

The due date for submitting this assignment has passed.

Due on 2022-03-23, 23:59 IST.

Assignment submitted on 2022-03-23, 07:31 IST

1) For categorical data with 'n' categories, the number of dummy variables will be_____ ***1 point***

- n
- n - 1
- n + 1
- 2n

Yes, the answer is correct.
Score: 1
Accepted Answers:
n - 1

2) In the estimation of regression parameters ***1 point***

- The likelihood function is a function of only σ
- The values of $\beta 0..\beta n$ and σ should be such that, they maximize the likelihood function.
- Both a. and b.
- None of these

No, the answer is incorrect.
Score: 0
Accepted Answers:
The values of $\beta 0..\beta n$ and σ should be such that, they maximize the likelihood function.

3) In logistic regression, the null hypothesis tested is: ***1 point***

- H0: β = 0

- $H0: \beta \neq 0$
- H0: μ = 0
- $H0: \mu \neq 0$

Yes, the answer is correct.
Score: 1
Accepted Answers:
H0: β = 0

4) In logistic regression, ***1 point***

- The graph doesn't follow S shape curve
- The dependent variable is categorical
- The estimated value of the dependent variable is not probability
- None of these

Yes, the answer is correct.
Score: 1
Accepted Answers:
The dependent variable is categorical

5) State true or false: G statistic is used to check the individual significance of the independent variables ***1 point***

- True
- False

Yes, the answer is correct.
Score: 1
Accepted Answers:
False

6) Choose the correct statement ***1 point***

- In logistic regression, the dependent variable must be continuous data
- In logistic regression, the dependent variable must be categorical data
- In logistic regression, both dependent and independent variables must be categorical data
- None of these

No, the answer is incorrect.
Score: 0
Accepted Answers:
In logistic regression, the dependent variable must be categorical data

7) State True or False: The Method of Least Squares can be applied to models with any probability distribution. ***1 point***

- True
- False

Yes, the answer is correct.
Score: 1
Accepted Answers:
False

8) Suppose you have been given a fair coin and you want to find out the odds of getting heads. Which of the following option is true for such a ***1 point***

Solve sample problems with us ()

case?

- Odds will be 0
- Odds will be 0.5
- Odds will be 1
- None of these

Yes, the answer is correct.
Score: 1

Accepted Answers:
Odds will be 1

9) Large values of the log-likelihood statistic indicate: ***1 point***

- That there are a greater number of explained vs. unexplained observations.
- That the statistical model fits the data well.
- That as the predictor variable increases, the likelihood of the outcome occurring decreases.
- That the statistical model is a poor fit of the data

No, the answer is incorrect.
Score: 0

Accepted Answers:
That the statistical model is a poor fit of the data

10) The logit function(given as l(x)) is the log of odds function. What could be the range of logit function in the domain x=[0,1]? ***1 point***

- (– ∞ , ∞)
- (0,1)
- (0 , ∞)
- (- ∞, 0)

Yes, the answer is correct.
Score: 1

Accepted Answers:
(– ∞ , ∞)

X

Week9 : Assignment 9

The due date for submitting this assignment has passed.

Due on 2022-03-30, 23:59 IST.

Assignment submitted on 2022-03-30, 19:40 IST

1) State true or false: Statement: there is no difference between, $y = \beta0 + \beta1x + \epsilon$ and $E(y) = \beta0 + \beta1x$, both are regression equations ***1 point***

- (●) True
- ( ) False

No, the answer is incorrect.
Score: 0

Accepted Answers:
False

2) Which of the following statements is correct ***1 point***

- (●) Sensitivity in ROC analysis is called True Positive Rate(tpr)
- ( ) Specificity in ROC analysis is not called True Negative Rate (tnr)
- ( ) Specificity in ROC analysis is called True Positive Rate(tpr)
- ( ) Sensitivity in ROC analysis is called True Negative Rate (tnr)

Yes, the answer is correct.
Score: 1

Accepted Answers:
Sensitivity in ROC analysis is called True Positive Rate(tpr)

3) In ROC analysis when the Threshold value is Higher: ***1 point***

- ( ) Specificity decreases
- (●) Sensitivity decreases
- ( ) Both a. and b.
- ( ) None of these

Yes, the answer is correct.
Score: 1

Accepted Answers:
Sensitivity decreases

4) Sensitivity in ROC analysis is defined as: ***1 point***

- FP / (FP+TN)
- FN/(TP+FN)
- TN / (TN+FP)
- TP / (TP+FN)

Yes, the answer is correct.
Score: 1

Accepted Answers:
TP / (TP+FN)

5) In ROC analysis, a classifier is called 'good' if it has _____ ***1 point***

- Low TPR and Low FPR
- Low TPR and High FPR
- High TPR and Low FPR
- High TPR and High FPR

Yes, the answer is correct.
Score: 1

Accepted Answers:
High TPR and Low FPR

6) For the given confusion matrix, compute the sensitivity ***1 point***

	True Positive	True Negative
Predicted Positive	8	3
Predicted Negative	2	7

- 0.73
- 0.7
- 0.78
- 0.8

Yes, the answer is correct.
Score: 1

Accepted Answers:
0.8

7) State true or False: Precision is inversely proportional to sensitivity ***1 point***

- True
- False

Yes, the answer is correct.
Score: 1

Accepted Answers:
False

8) State True or False: Standardization of features is not required before training a Logistic regression model ***1 point***

- True

False

Yes, the answer is correct.
Score: 1

Accepted Answers:
True

9) Which of the following option is true? ***1 point***

- Linear Regression errors values have to be normally distributed but in the case of Logistic Regression it is not the case
- Logistic Regression errors values have to be normally distributed but in the case of Linear Regression it is not the case
- Both Linear Regression and Logistic Regression error values have to be normally distributed
- Both Linear Regression and Logistic Regression error values have not to be normally distributed

Yes, the answer is correct.
Score: 1

Accepted Answers:
Linear Regression errors values have to be normally distributed but in the case of Logistic Regression it is not the case

10) In binary logistic regression, ***1 point***

- The dependent variable is continuous
- The dependent variable is divided into two equal subcategories
- The dependent variable consists of two categories
- There is no dependent variable

Yes, the answer is correct.
Score: 1

Accepted Answers:
The dependent variable consists of two categories

X

Week10 : Assignment 10

The due date for submitting this assignment has passed.

Due on 2022-04-06, 23:59 IST.

Assignment submitted on 2022-04-06, 19:58 IST

1) Sampling distribution for the goodness of fit test is the ***1 point***

- Poisson distribution
- t distribution
- normal distribution
- chi-square distribution

Yes, the answer is correct.
Score: 1

Accepted Answers:
chi-square distribution

2) The goodness of fit test is always conducted as a ***1 point***

- lower-tail test
- upper-tail test
- middle test
- None of these

Yes, the answer is correct.
Score: 1

Accepted Answers:
upper-tail test

3) State True or False: Statement: Null hypothesis for the chi-square test of independence assumes that all the proportions are equal. ***1 point***

- True
- False

Yes, the answer is correct.
Score: 1

Accepted Answers:
True

4) Statistical test conducted to determine whether to reject or not reject a hypothesized probability distribution for a population is known as a _____ ***1 point***

- contingency test
- probability test
- goodness of fit test
- None of these

Yes, the answer is correct.
Score: 1

Accepted Answers:
goodness of fit test

5) What is the minimum no. of variables/ features required to perform clustering? ***1 point***

- 0
- 1
- 2
- 3

Yes, the answer is correct.
Score: 1

Accepted Answers:
1

6) State True or False: The chi-square test of independence is used to analyze the frequencies of two variables with multiple categories and determine whether they are independent ***1 point***

- True
- False

Yes, the answer is correct.
Score: 1

Accepted Answers:
True

7) State True or False: Minkowski distance is a generalization of both Euclidean and Manhattan metrics ***1 point***

- True
- False

Yes, the answer is correct.
Score: 1

Accepted Answers:
True

8) In order to perform the chi-square test to check independence in python using scipy, we import _____ ***1 point***

- chi2_test
- chi_square_independence
- chi2_contingency
- None of these

Yes, the answer is correct.
Score: 1

Accepted Answers:
chi2_contingency

9) Let x1 = (1,2) and x2 = (3,5) be the co-ordinates for two objects. The Euclidean and Manhattan distance between these two objects is _____ respectively **1 point**

- 4.2 and 3
- 3.15 and 2
- 3.61 and 5
- None of these

Yes, the answer is correct.
Score: 1

Accepted Answers:
3.61 and 5

10) State true or false: Discriminant Analysis does not require the grouping variable to be known at the beginning **1 point**

- True
- False

Yes, the answer is correct.
Score: 1

Accepted Answers:
False

X


(https://swayam.gov.in)  (https://swayam.gov.in/nc_details/NPTEL)

Week11 : Assignment 11

The due date for submitting this assignment has passed.

Due on 2022-04-13, 23:59 IST.

Assignment submitted on 2022-04-13, 22:23 IST

1) _____ is used for calculating distance measures in clustering using python *1 point*

- (●) distance_matrix
- ( ) spatial_matrix
- ( ) scipy_matrix
- ( ) distance.matrix

Yes, the answer is correct.
Score: 1

Accepted Answers:
distance_matrix

2) The formula for dissimilarity computation between two objects for categorical variables is – *1 point*

Here p is a categorical variable and m denotes the number of matches.

- (●) D(i,j) = p-m / p
- ( ) D(i,j) = p-m / m
- ( ) D(i,j) = m-p / p
- ( ) D(i,j) = m-p / m

Yes, the answer is correct.
Score: 1

Accepted Answers:
D(i,j) = p-m / p

3) Select the correct option for a data set with 7 objects and an interval-scaled variable ‘f’ we have the following measurements: f = (1, 2, 3, 4, 5, 8, 50) containing one outlying value. ***1 point***

- Std deviation (std_f) and mean absolute deviation (s_f) are equally affected
- Mean absolute deviation (s_f) is more affected by the outlier
- Std deviation (std_f) is more affected by the outlier
- None of these

Yes, the answer is correct.
Score: 1
Accepted Answers:
Std deviation (std_f) is more affected by the outlier

4) Which of the following is true for K-means clustering? ***1 point***

- It comes under the partitioning method
- The number of clusters is predefined for this method
- Cluster similarity is measure in regard to the mean value of the objects in a cluster
- All of the above

Yes, the answer is correct.
Score: 1
Accepted Answers:
All of the above

5) Which of the following can act as possible termination conditions in K-Means? ***1 point***

1. For a fixed number of iterations.
2. Assignment of observations to clusters does not change between iterations. Except for cases with a bad local minimum.
3. Centroids do not change between successive iterations.
4. Terminate when Residual Sum of Squares (RSS) falls below a threshold.

- 1,3 and 4
- 1,2,3 and 4
- 2 and 3
- None of these

Yes, the answer is correct.
Score: 1
Accepted Answers:
1,2,3 and 4

6) In the figure below, if you draw a horizontal line on y-axis for y=2. What will be the number of clusters formed? ***1 point***

(Image link: https://drive.google.com/file/d/1uLZZgkWh7SwYIPg9S5PSwHc3kqGAIJbI/view)

- 1
- 2
- 3
- 4

Yes, the answer is correct.
Score: 1

Accepted Answers:
2

7) Which of the following clustering requires merging approach? ***1 point***

- Partitional
- Naive Bayes
- Hierarchical
- None of these

Yes, the answer is correct.
Score: 1
Accepted Answers:
Hierarchical

8) State True or False: Hierarchical clustering should primarily be used for exploration ***1 point***

- True
- False

Yes, the answer is correct.
Score: 1
Accepted Answers:
True

9) State True or False: For finding dissimilarity between two clusters in hierarchical clustering, average-link is the only metric used ***1 point***

- True
- False

Yes, the answer is correct.
Score: 1
Accepted Answers:
False

10) Hierarchical clustering can either be an agglomerative or divisive algorithm ***1 point***

- True
- False

Yes, the answer is correct.
Score: 1
Accepted Answers:
True