

Sample Questions

Circle the right answers

Question: What does ETL stands for

Answer: Extract, Transform and Load

Question: A Person correlation between two random variables is zero, but their values can still be related to each other (True or False)

- A. True
- B. False

Answer: A

Question: Explain two functional differences between a Database vs a Data Warehouse

Answer:

Question: What is a categorical variable?

Answer:

Question: Name two common methods in filling up missing values for time-series data

Answer: forward and backward

Question: Which of the following statements is/are true about Type-1 and Type-2 errors?

1. Type-1 error occurs when the prediction is different from the actual case
2. Type-2 error is when the prediction is negative while the actual case is positive
3. Type-2 occurs when we fail to accept a null hypothesis when it is actually true

- A. Only 1
- B. Only 2
- C. 1 and 2
- D. 1 and 3
- E. None of the above

Answer: B , 3 is false positive

Question:

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

- A. Calculate the Precision: Your answer: $TP/(TP+FP) = 100/110$
- B. Calculate the Recall: Your answer: $TP/(TP+FN) = 100/105$
- C. Calculate the True Positive Rate: answer: $TP/(TP+FN) = \text{Recall}$
- D. Calculate the False Positive Rate: answer: $FP/(FP+TN) = 10/(10+50)$

Fill in the blank for the following questions

Question: The height distribution taken from Computer Science Major in Queen College will have a mean ____ (higher, lower or similar) than the whole college and a ____ (positive, zero or negative) skews

Question: The height distribution taken from the basketball team in Queen College will have a mean ____ (higher, lower or similar) than the whole college and a ____ (positive, zero or negative) skews

Question: The height distribution taken from Computer Science class in Queen College will have a mean ____ (higher, lower or similar) than the whole college and ____ (positive/zero/negative) skews if we know many are also in the basketball Team

Question: The household income distribution of a gated community should have a ____ (higher/same/lower) standard deviation than a random sample of the whole population

Question: Given the following database tables, Write a SQL statement to prepare a list with salesman name, customer name and their cities for the salesmen and customer who belongs to the same city.

Sample table: salesman

salesman_id	name	city	commission
5001	James Hoog	New York	0.15
5002	Nail Knite	Paris	0.13
5005	Pit Alex	London	0.11
5006	Mc Lyon	Paris	0.14
5007	Paul Adam	Rome	0.13
5003	Lauson Hen	San Jose	0.12

Sample table: customer

customer_id	cust_name	city	grade	salesman_id
3002	Nick Rimando	New York	100	5001
3007	Brad Davis	New York	200	5001
3005	Graham Zusi	California	200	5002
3008	Julian Green	London	300	5002
3004	Fabian Johnson	Paris	300	5006
3009	Geoff Cameron	Berlin	100	5003
3003	Jozv Altidor	Moscow	200	5007

Answer:

```
SELECT salesman.name AS "Salesman", customer.cust_name, customer.city FROM salesman, customer WHERE salesman.city=customer.city;
```

Question: Using the previous tables, Write a SQL statement to find the list of customers who appointed a salesman for their jobs who gets a commission from the company is more than 12%.

```
Answer: SELECT a.cust_name AS "Customer Name", a.city, b.name AS "Salesman", b.commission FROM customer a INNER JOIN salesman b ON a.salesman_id=b.salesman_id WHERE b.commission>.12;
```

Question: Which of the following are true

1. In a positively skewed distribution, the mean will be less than the median
2. In a negatively skewed distribution, the median will be larger than the mean
3. In a normal distribution, the mean and the mode are the same

- A. Only 1
- B. Only 2
- C. Only 3
- D. 1 and 3
- E. 2 and 3

Answer:

Question: Suppose you are given the following plots 1-4 (from left to right) and you want to compare their Pearson correlation coefficients. Which of the following are true (including the sign)?



1. $1 > 2$
2. $3 > 4$
3. $3 > 2$

- A. Only 1
B. Only 2
C. 1 and 2
D. 2 and 3
E. 1, 2, and 3

Answer: C

Question: What is the difference between random sampling vs systematic sampling?

Answer: