Feedback — Week 2 Quiz

Help Center

Thank you. Your submission for this quiz was received.

You submitted this quiz on **Wed 18 Feb 2015 8:42 AM PST**. You got a score of **11.00** out of **11.00**.

Question 1

Suppose a school collected some data on students' preference for hot dogs(HD) vs. hamburgers(HM). We have the following 2×2 contingency table summarizing the statistics. If lift is used to measure the correlation between HD and HM, what is the value for lift(HD, HM)?

	HD	¬HD	Σ_{row}
HM	40	24	64
¬HM	210	126	336
Σ_{col}	250	150	400

Your Answer		Score	Explanation
◉ 1	✔	1.00	
○ 0			
○ -1			
○ -∞			
Total		1.00 / 1.00	

Question Explanation

The correct answer is: "1".

The lift can be calculated by

$$\text{Lift} = \frac{\text{supp}(A \cup B)}{\text{supp}(A) \cdot \text{supp}(B)},$$

where supp(A) and supp(B) refer to the relative support of A and B respectively. Thus,

$$\text{Lift} = \frac{40/400}{250/400 \times 64/400} = 1.$$

Question 2

Suppose a school collected some data on students' preference for hot dogs(HD) vs. hamburgers(HM). We have the following 2×2 contingency table summarizing the statistics. If χ^2 is used to measure the correlation between HD and HM, what is the χ^2 score?

	HD	¬HD	Σ_{row}
HM	40	24	64
¬HM	210	126	336
Σ_{col}	250	150	400

Your Answer		Score	Explanation
-∞			
1			
-1			
0	✔	1.00	
Total		1.00 / 1.00	

Question Explanation

The correct answer is: "0".

The contingency table with expected values is following table

	HD	¬HD	Σ_{row}
HM	40 (40)	24 (24)	64
¬HM	210 (210)	126 (126)	336
Σ_{col}	250	150	400

χ^2 can be evaluated as follows

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the observed frequency, and E_i is the expected frequency. Since the expected values equal the observed ones, we have $\chi^2 = 0$.

Question 3

What is the value range of the Kulczynski measure?

Your Answer		Score	Explanation
◉ [0, 1]	✔	1.00	
○ [-1, 1]			
○ [0, +∞)			
○ (-∞, +∞)			
Total		1.00 / 1.00	

Question Explanation

By definition, the correct answer is: "[0, 1]".

Question 4

Which of the following measures is NOT null invariant?

Your Answer		Score	Explanation
○ Kulcyzynski			
○ Cosine			
◉ χ^2	✔	1.00	
○ All confidence			
Total		1.00 / 1.00	

Question Explanation

The correct answer is: "χ^2".

Null transactions are considered in χ^2.

Question 5

Suppose we are interested in analyzing the transaction history of several supermarkets with respect to purchase of apples(A) and bananas(B). We have the following table summarizing the transactions.

Supermarket	AB	¬AB	A ¬B	¬A ¬B
S1	100,000	7,000	3,000	300
S2	100,000	7,000	3,000	90,000

Denote l_i as the lift measure and ki as the Kulcyzynski measure for supermarket S_i (i = 1, 2). Which of the following is correct?

Your Answer		Score	Explanation
○ $l_1 = l_2$, $k_1 = k_2$			
○ $l_1 = l_2$, $k_1 \neq k_2$			
○ $l_1 \neq l_2$, $k_1 \neq k_2$			
◉ $l_1 \neq l_2$, $k_1 = k_2$	✔	1.00	
Total		1.00 / 1.00	

Question Explanation

The correct answer is: "$l_1 \neq l_2$, $k_1 = k_2$".

Lift is not null invariant and therefore sensitive to the number of null transactions, while Kulcyzynski is null invariant.

Question 6

A store had 100,000 total transactions in Q4 2014. 10,000 transactions contained beer, while

5,000 contained frying pans. 600 transactions contained both beer and frying pans. Which of the following is true?

Your Answer		Score	Explanation
○ {beer, frying pans} is a negative pattern under the support-based definition of negatively correlated patterns.			
◉ For ε = 0.1, {beer, frying pans} is a negative pattern under the null-invariant definition of negatively correlated patterns.	✔	1.00	
○ There does not exists a value for ε such that {beer, frying pans} is a negative pattern by the null-invariant definition of negative patterns.			
○ More information is needed to determine if {beer, frying pans} is a negative pattern.			
Total		1.00 / 1.00	

Question Explanation

The correct answer is "For ε = 0.1, {beer, frying pans} is a negative pattern under the null-invariant definition of negatively correlated patterns."

Support based:

$$\text{sup}(\text{beer}) = \frac{10000}{100000} = 0.1$$
$$\text{sup}(\text{frying pans}) = \frac{5000}{100000} = 0.05$$
$$\text{sup}(\text{beer} \cup \text{frying pans}) = \frac{600}{100000} = 0.006$$
$$\text{sup}(\text{beer}) * \text{sup}(\text{frying pans}) = 0.1 * 0.05 = 0.005 < \text{sup}(\text{beer} \cup \text{frying pans})$$

Null-invariant:

$$\frac{P(\text{beer}|\text{frying pans}) + P(\text{frying pans}|\text{beer})}{2} = \frac{600/5000 + 600/10{,}000}{2} = 0.09$$

Thus, {beer, frying pans} is not a negative pattern by the support-based definition. For the null-invariant definition, it is only considered a negative pattern for ε = 0.1.

Question 7

Pat-ID	Item-Sets	Support
P1	{A, C, E, S}	205227
P2	{F, A, C, E, S}	205211
P3	{F, A, C, E, T, S}	101758
P4	{F, A, C, T, S}	161563
P5	{A, C, T, S }	161576

Table 1: Support for frequent itemsets

Given the itemsets in Table 1, which of the following patterns are in the δ-cluster containing the pattern {A, C, E, S} for δ = 0.0001?

Your Answer		Score	Explanation
◉ {F, A, C, E, S}	✔	1.00	
○ {F, A, C, T, S}			
○ {F, A, C, E, T, S}			
○ {A, C, T, S}			
Total		1.00 / 1.00	

Question Explanation

The correct answer is: "{F, A, C, E, S}".

All the other patterns have a distance greater than δ from the pattern {A, C, E, S}.

Question 8

Transactions	# of Transactions
(abe)	100
(bcf)	100
(acf)	100
(abcef)	100

Table 2: # transactions in a database.

Given the transactions in Table 2, which of the following is a 0.5-core pattern of (abcef)? Select all that apply

Your Answer	Score	Explanation

☑ (acfe)	✔	0.20
☐ (a)	✔	0.20
☑ (e)	✔	0.20
☑ (abcef)	✔	0.20
☑ (abe)	✔	0.20
Total		1.00 / 1.00

Question Explanation

The correct answers are: "(acfe)", "(abcef)", "(e)", and "(abe)".

Every pattern except (a) is contained in at most one other pattern in the database, making it a 0.5-core pattern of (abcef).

Question 9

A constraint is anti-monotone if an itemset S violates the constraint, so do all of its supersets.

Which of following constraints is anti-monotone?

Your Answer		Score	Explanation
○ var(S.price) > 20			
○ sum(S.price) > 25			
○ avg(S.profit) > 15			
◉ range(S.profit) < 10	✔	1.00	
Total		1.00 / 1.00	

Question Explanation

The correct answer is: "range(S.profit) < 10"

Suppose X is the superset of S; if so, the following inequalities always hold for all X:

- avg(X.price) <> avg(S.price)
- sum(X.price) ≥ sum(S.price)
- var(X.profit) <> var(S.profit)
- range(X.profit) ≥ range(S.profit)

From these inequalities, we can easily verify that range(S.profit) < 10 is anti-monotone.

Question 10

A constraint is monotone if an itemset S satisfies the constraint, so do all of its supersets. Which of following constraints is monotone?

Your Answer		Score	Explanation
○ Support of S > 100			
○ avg(S.price) < 10			
◉ sum(S.price) > 20	✔	1.00	
○ min(S.price) > 10			
Total		1.00 / 1.00	

Question Explanation

The correct answer is: "sum(S.price) > 20"

Suppose X is the superset of S; if so, the following inequalities always hold for all X:

- support of X ≤ support of S
- min(X.price) ≤ min(S.price)
- avg(X.price) <> avg(S.price)
- sum(X.price) ≥ sum(S.price)

From these inequalities, we can easily verify that sum(S.price) > 20 is monotone.

Question 11

A constraint is succinct if the constraint c can be enforced by directly manipulating the data. Which of following constraints is NOT succinct?

Your Answer		Score	Explanation
○ min(S.profit) < 40			
○ v ∈ S			
○ min(S.profit) > 40			
◉	✔	1.00	It cannot be determined beforehand since range of the price of

range(S.price) > 2		itemset S keeps increasing.
Total	1.00 / 1.00	

Question Explanation

The correct answer is: "range(S.price) > 2"