

Interestingness Measure: Lift

Lift is more telling than s & c

¬B

350

50

400

 Σ_{row}

750

250

1000

В

400

200

600

C

-C

Measure of dependent/correlated events: lift

$$lift(B,C) = \frac{c(B \to C)}{s(C)} = \frac{s(B \cup C)}{s(B) \times s(C)}$$

☐ Lift(B, C) may tell how B and C are correlated

□ Lift(B, C) = 1: B and C are independent

□ > 1: positively correlated

< 1: negatively correlated</p>

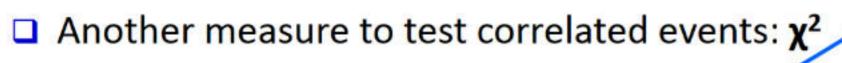
For our example,
$$lift(B,C) = \frac{400/1000}{600/1000 \times 750/1000} = 0.89$$

$$lift(B,\neg C) = \frac{200/1000}{600/1000 \times 250/1000} = 1.33$$

- □ Thus, B and C are negatively correlated since lift(B, C) < 1;</p>
 - B and ¬C are positively correlated since lift(B, ¬C) > 1

2

Interestingness Measure: χ^2



$u^2 - \nabla (Ob)$	$served - Expected)^2$
$\chi - Z$	Expected

	_	V COLUMN TO SERVICE	posters 4	Service and the
ш.	Ger	neral	ru	es

- \square $\chi^2 = 0$: independent
- χ² > 0: correlated, either positive or negative, so it needs additional test

				10
		В	¬ B	Σ_{row}
C	7740	00 (450)	350 (300)	750
¬C	2	J (150)	50 (100)	250
Σ_{col}		600	400	1000

Expected value

Observed value

Now,
$$\chi^2 = \frac{(400 - 450)^2}{450} + \frac{(350 - 300)^2}{300} + \frac{(200 - 150)^2}{150} + \frac{(50 - 100)^2}{100} = 55.56$$

- χ² shows B and C are negatively correlated since the expected value is 450 but the observed is only 400
- \square χ^2 is also more telling than the support-confidence framework

Lift and χ^2 : Are They Always Good Measures?

- Null transactions: Transactions that contain neither B nor C
- Let's examine the dataset D
 - BC (100) is much rarer than B¬C (1000) and ¬BC (1000), but there are many ¬B¬C (100000)
 - Unlikely B & C will happen together!
- But, Lift(B, C) = 8.44 >> 1 (Lift shows B and C are strongly positively correlated!)
- \square χ^2 = 670: Observed(BC) >> expected value (11.85)
- ☐ Too many null transactions may "spoil the soup"!

1	В	¬B	Σ_{row}
С	100	1000	1100
¬C	1000	100000	101000
$\Sigma_{\text{col.}}$	1100	101000	102100
		null transactions	

Contingency table with expected values added

	В	¬B	Σ_{row}
С	100 (11.85)	1000	1100
¬C	1000 (988.15)	100000	101000
$\Sigma_{\text{col.}}$	1100	101000	102100