Generalized form of the N by M contingency test with expected values based on joint products of observed values

Colin James III, PhD

Abstract

The general form is described for an N by M contingency test that uses joint products of observed values to produce the expected values.

Introduction

Deriving expected values from manipulation of observed values may be used for contingency tests, of which the Chi-squared test is a subset. Using joint products is more accurate than using group sum totals as shown by James, 1981.

Consider this 3 by 3 contingency table:

	x	y	z
p	a1	b1	n1
q	b2	a2	n2
r	c1	c2	n3

with row totals:

p	=	a1 + b1 + n1
q	=	b2 + a2 + n2
r	=	c1 + c2 + n3

and column totals:

x	=	a1 + b2 + c1
y	=	b1 + a2 + c2
z	=	n1 + n2 + n3

The $\chi 2$ statistic (Chi-squared) based on observed and expected values derived from the observed values by joint products is calculated here:

```
( p + q + r) *
(
  ( a1 ^ 2) / ( p * x)
+ ( b1 ^ 2) / ( p * y)
+ ( n1 ^ 2) / ( p * z)
+ ( b2 ^ 2) / ( q * x)
+ ( a2 ^ 2) / ( q * y)
+ ( n2 ^ 2) / ( q * z)
+ ( c1 ^ 2) / ( r * x)
+ ( c2 ^ 2) / ( r * y)
+ ( n3 ^ 2) / ( r * z)
- 1
)
```

where df = degrees of freedom = (rows – 1) * (columns – 1) .

Generalization

The generalized form is intuitive from the table above.

	x	y … column N
p	a1	b1 … n1
q	b2	a2 … n2
:	:	: :
row M	m1	m2 … NM

with row totals:

p	=	a1 + b1 + … + n1
q	=	b2 + a2 + … + n2
:	:	:
M	=	c1 + c2 + … + NM

and column totals:

x	=	a1 + b2 + … + m1
y	=	b1 + a2 + … + m2
:	:	:
N	=	n1 + n2 + … + NM

χ^2 =

```
( p + q + ... + M) *
(
  ( a1 ^ 2) / ( p * x)
+ ( b1 ^ 2) / ( p * y)
+ ...
+ ( n1 ^ 2) / ( p * N)
+ ( b2 ^ 2) / ( q * x)
+ ( a2 ^ 2) / ( q * y)
+ ...
+ ( n2 ^ 2) / ( q * N)
+ ...
+ ( m1 ^ 2) / ( M * N)
- 1
)
```

where df = (M – 1) * (N – 1) .

Acknowledgments

Thanks are due for helpful discussion with Geoffrey A Dean.

References

James, C, 1981, CAS Technical Paper No 15. 1981. Unpublished.