ORIGIN ≡ 1

χ^2 & G Tests for Association in RXC Contingency Tables

This test employs a RXC contingency table consisting of R rows and C Columns and is thus an extension of the 2X2 case discussed in *Biostatistics* 440.

Data Structure:

Row by Column table of counts.

RXC Contingency Table					
$O_{1,1}$	$O_{1,2}$	$O_{1,3}$	…	$O_{1,j}$	
$O_{2,1}$	$O_{2,2}$	$O_{2,3}$	…	$O_{2,j}$	
$O_{3,1}$	$O_{3,2}$	$O_{3,3}$	…	$O_{3,j}$	Row Totals
…	…	…	…	…	
$O_{i,1}$	$O_{i,2}$	$O_{i,3}$	…	$O_{i,j}$	
	Column Totals				Grand Total

Assumptions:

- **Observed values $X_{i,j}$ are a random sample**
- **Observed values for Rows and Columns are independent.**

Model:

Let Probabilities:

- **$P_i = P(R=i)$**
- **$P_j = P(C=j)$**
- **$P_{ij} = P(R=i, C=j)$**

Zar Example 23.1

	Hair Color				
Sex	*black*	*brown*	*blond*	*red*	*total*
male	**32**	**43**	**16**	**9**	100
female	**55**	**65**	**64**	**16**	200
total	87	108	80	25	300

$$O := \begin{pmatrix} 32 & 43 & 16 & 9 \\ 55 & 65 & 64 & 16 \end{pmatrix}$$

$R := 2$ **< rows**

$C := 4$ **< columns**

Hypotheses:

H_0: $P_{ij} = (P_i)(P_j)$ **< That is, variables R & C are independent!**

H_1: $P_{ij} <> (P_i)(P_j)$ **< Two sided test**

Criterion for Normal Approximation:

- **IF no more than 1/5 of the cells have expected values in each block** $E_{ij} \le 5$ **AND no cell has expected value** $E_{ij} < 1$ **THEN Approximation may be used**

Construct Contingency Tables of Observed and Expected in each cell:

- **Tabulate O_{ij} for each cell**
- **Calculate Observed Row and Column Totals**
- **Calculate Expected for each cell**

$$E := \begin{pmatrix} \frac{87 \cdot 100}{300} & \frac{108 \cdot 100}{300} & \frac{80 \cdot 100}{300} & \frac{25 \cdot 100}{300} \\ \frac{87 \cdot 200}{300} & \frac{108 \cdot 200}{300} & \frac{80 \cdot 200}{300} & \frac{25 \cdot 200}{300} \end{pmatrix} \qquad E = \begin{pmatrix} 29 & 36 & 26.6667 & 8.3333 \\ 58 & 72 & 53.3333 & 16.6667 \end{pmatrix}$$

Test Statistics:

$i := 1..R$ $j := 1..C$

$$\chi := \sum_i \sum_j \frac{\left(O_{i,j} - E_{i,j}\right)^2}{E_{i,j}} \qquad \chi = 8.9872$$

< chi-square statistic

$$G := 2 \cdot \left(\sum_i \sum_j O_{i,j} \cdot \ln\left(\frac{O_{i,j}}{E_{i,j}}\right)\right) \qquad G = 9.5121$$

< log-liklihood g statistic

Sampling Distribution of Test Statistics:

If Assumptions hold and H_0 is true then: $\chi \sim \chi^2_{((R-1)(C-1))}$

$G \sim \chi^2_{((R-1)(C-1))}$

< where R & C are the number of Row and Column cells respectively

Critical Value of the Test:

$\alpha := 0.05$ **< Probability of Type I error must be explicitly set**

$df := (R-1)\cdot(C-1)$ $df = 3$

$CV := qchisq(1-\alpha, df)$ $CV = 7.8147$

Decision Rule:

IF χ > CV THEN REJECT $H_{0,}$ OTHERWISE ACCEPT H_0
IF G > CV THEN REJECT $H_{0,}$ OTHERWISE ACCEPT H_0

Probability Value:

$P_\chi := (1 - pchisq(\chi, df))$ $P_\chi = 0.02946$

$P_G := (1 - pchisq(G, df))$ $P_G = 0.0232$

Prototype in R:

```
#RXC CONTINGENCY TEST
#ZAR EXAMPLE 23.1

#DATA KEYED INTO MATRIX:
X=matrix(c(32,43,16,9,55,65,64,16),nrow=2,byrow=T)
X

chisq.test(X)
```

```
        Pearson's Chi-squared test

data:  X
X-squared = 8.9872, df = 3, p-value = 0.02946
```

```
#HAND CALCULATION OF G TEST:
O=matrix(chisq.test(X)$observed)
O
E=matrix(chisq.test(X)$expected)
E
```

< these commands in R extract previously Observed and Expected values from the output of chisq.test() and puts them into vectors O & E using matrix()

```
#G STATISTIC:
G=2*sum(O*log(O/E))
G
```

```
> G
[1] 9.512149
```

```
#CRITICAL VALUE:
alpha=0.05
R=2
C=4
df=(R-1)*(C-1)
CV=qchisq(1-alpha,df)
CV
```

```
> C
[1] 7.814728
```

```
#PROBABILITY:
P=(1-pchisq(G,df))
P
```

```
> P
[1] 0.02320247
```

```
#DATA FROM A (DIFFERENT) ORIGINAL DATASET IN R FORMAT:
ROS=read.table("c:/DATA/Biostatistics/RosnerTA3.11.txt")
ROS
attach(ROS)
```

```
> ROS
  Clear Antibo Age Ear
1     1      1   1   1
2     1      1   1   1
3     1      1   1   1
4     0      1   1   1
5     0      1   1   1
6     0      1   1   1
… MANY MORE…
```

```
#CREATING CONTINGENCY TABLE:
T=table(Antibo,Age)
T
```

```
> T
      Age
Antibo  1  2  3
     1 49 74 27
     2 42 54 32
```

```
chisq.test(T)
```

```
        Pearson's Chi-squared test

data:  T
X-squared = 2.361, df = 2, p-value = 0.3071
```