ORIGIN ≡ 1

2 X 2 Contingency Test of Association

Contingency tests consider data from categorical (also called nominal) variables - variables in which observations may be placed in classes, but the classes themselves need not have numerical or ordinal significance. When comparing two categorical variables it is customary to construct a contingency table showing which observations may be simultaneously classified according to the classes. From the contingency table, tests of association (or alternatively tests of independence) may be performed. Here we look at the 2X2 case in which there are only 2 classes for each of two variables.

Data Structure:

Zar 2010 pp. 497-499 describes three different interpretations of a 2X2 contingency table depending on what's determined in advance by the researcher:

1) *No fixed margins* - row and column totals are not determined in advance.

2) *One fixed margin* - row or column totals determined in advance but not both.

3) *Both margins fixed* - both row and column totals are determined in advance

Disagreements exist over whether or not to correct the test statistic accordingly. Zar suggests use of Yate's correction shown below in the case of 3) only. Other correction methods are also available.

Contingency Table		
1,1	1,2	X=1
2,1	2,2	X=2
Y=1	Y=2	Grand Total

Assumptions:

- Observed values $X_1, X_2, X_3, \ldots X_{n1}$ are a random sample
Observed values $Y_1, Y_2, Y_3, \ldots Y_{n2}$ are a random sample.

Model:

Let Probabilities:
- $P_i = P(X=i)$
- $P_j = P(Y=j)$
- $P_{ij} = (X=i, Y=j)$

$O := \begin{pmatrix} 6 & 12 \\ 28 & 24 \end{pmatrix}$ **< observed**

$GM := 70$ **< grand total**

Zar Example 23.2

	boys	girls	*total*
left-handed	**6**	**12**	18
right-handed	**28**	**24**	52
total	34	36	70

Hypotheses:

H_0: $P_{ij} = (P_i)(P_j)$ **< That is, variables X & Y are independent!**
H_1: $P_{ij} <> (P_i)(P_j)$ **< Two sided test**

Criterion for Normal Approximation:

- IF expected values in each cell $E_{ij} \geq 5$ THEN Normal Approximation may be used
OTHERWISE use an Exact Test such as Fisher's Exact Test

Normal Approximation:

Construct Expected for each cell:

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

$$E := \begin{pmatrix} \frac{18 \cdot 34}{70} & \frac{18 \cdot 36}{70} \\ \frac{52 \cdot 34}{70} & \frac{52 \cdot 36}{70} \end{pmatrix} \qquad E = \begin{pmatrix} 8.7429 & 9.2571 \\ 25.2571 & 26.7429 \end{pmatrix}$$

^ multiply row total by column total and divide by total

Test Statistic:

$i := 1..2 \qquad j := 1..2$

$$\chi := \sum_i \sum_j \frac{\left(O_{i,j} - E_{i,j}\right)^2}{E_{i,j}} \qquad \chi = 2.2524$$ **< Standard statistic**

$$\chi_Y := \sum_i \sum_j \frac{\left(\left|O_{i,j} - E_{i,j}\right| - \frac{1}{2}\right)^2}{E_{i,j}} \qquad \chi_Y = 1.5061$$ **< Yate's corrected statistic**

Sampling Distribution:

If Assumptions hold, H_0 is true and assuming Normal Approximation, then $\chi_Y \sim \chi^2_{(1)}$

Critical Value of the Test:

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

$CV := \text{qchisq}(1 - \alpha, 1) \qquad CV = 3.8415$

Decision Rule:

IF χ > CV THEN REJECT H_0, OTHERWISE ACCEPT H_0
IF χ_Y > CV THEN REJECT H_0, OTHERWISE ACCEPT H_0

Probability Value:

$P := (1 - \text{pchisq}(\chi, 1)) \qquad P = 0.1334$ **< Standard probability**

$P_Y := (1 - \text{pchisq}(\chi_Y, 1)) \qquad P_Y = 0.2197$ **< with Yate's corretion**

Prototype in R:

```
#2X2 CONTINGENCY TABLES
ZAR=read.table("c:/DATA/Biostatistics/ZarEX23.2R.txt")
ZAR
attach(ZAR)
Y=count
A=handedness
B=sex

#CROSS-TABULATION IN R:
X=xtabs(Y~A+B)
X

#USING CHI-SQUARE STATISTIC:
chisq.test(X,correct=F)
#USING YATE'S CORRECTED STATISTIC:
chisq.test(X,correct=T)

#DATA KEYED INTO MATRIX BY ROW
#USING CONCATENATION FUNCTION c()
AND MATRIX FUNCTION matrix():

X=matrix(c(6,12,28,24),nrow=2,byrow=T)

X
```

```
> X
      B
A       boys girls
  left     6    12
  right   28    24
```

```
        Pearson's Chi-squared test

data:  X
X-squared = 2.2524, df = 1, p-value = 0.1334


        Pearson's Chi-squared test with Yates' continuity correction

data:  X
X-squared = 1.5061, df = 1, p-value = 0.2197
```