Data frame and table

```
y <- c(320, 14, 80, 36)
particle <- gl(2, 1, 4, labels=c("no","yes"))
quality <- gl(2, 2, 4, labels=c("good","bad"))
(wafer <- data.frame(y, particle, quality))

##     y particle quality
## 1 320       no    good
## 2  14      yes    good
## 3  80       no     bad
## 4  36      yes     bad

(ov <- xtabs(y ~ quality + particle))
```

observed value y_{ij}
$i = 1, 2;\ j = 1, 2$

```
##        particle
## quality  no yes
##    good 320  14
##    bad   80  36
```

Marginal proportions

```
# multinomial model
# marginal proportions for particle values
(pp <- prop.table( xtabs(y ~ particle)))

## particle
##        no       yes
## 0.8888889 0.1111111

# marginal proportions for quality values
(qp <- prop.table( xtabs(y ~ quality)))

## quality
##      good       bad
## 0.7422222 0.2577778
```

$(\hat{\pi}_{\cdot j}\ ;\ j=1,2)$

$(\hat{\pi}_{i\cdot}\ ;\ i=1,2)$

Independence fitted values

$H_0 : \pi_{ij} = \pi_{i.} \times \pi_{.j}$

```
# multinomial model under independence
# # fitted values
(fv <- outer(qp,pp)*450)

##          particle
## quality        no      yes
##    good  296.8889 37.11111
##    bad   103.1111 12.88889

# deviance (on 1 d.f.)
2*sum(ov*log(ov/fv))

## [1] 54.03045

pchisq(54.03, 1, lower.tail=FALSE)

## [1] 1.974517e-13
```

$\hat{y}_{ij} = \hat{\pi}_{ij} y_{..} = \hat{\pi}_{i.} \times \hat{\pi}_{.j} \times y_{..}$

$$\text{Deviance} = -2 \log \left(\frac{\text{Likelihood under } H_0}{\text{Likelihood of full model}} \right)$$

Multinomial model

$$\Pr(Y_{ij} = y_{ij} \text{ for all } i \text{ and } j) = \frac{y_{..}!}{\prod_{ij} y_{ij}!} \prod_i \prod_j \pi_{ij}^{y_{ij}}$$

$$\Rightarrow \text{log likelihood of model} = c + \sum_i \sum_j y_{ij} \log \pi_{ij}$$

- Under H_0: $\hat{\pi}_{ij} = \hat{\pi}_{.j} \times \hat{\pi}_{i.} = \frac{y_{.j}}{y_{..}} \times \frac{y_{i.}}{y_{..}}$

- full model $\hat{\pi}_{ij} = \frac{y_{ij}}{y_{..}}$

$$\Rightarrow D = -2 \sum_i \sum_j (\text{Log likelihood } (H_0 \text{ model}) - \text{log likelihood (full model)})$$

$$= -2 \sum_i \sum_j y_{ij} \left(\log \hat{\pi}_{ij}(H_0) - \log \hat{\pi}_{ij}(\text{full}) \right)$$

$$= -2 \sum_i \sum_j y_{ij} \log \left(\hat{\pi}_{ij}(H_0) \times \frac{y_{..}}{y_{ij}} \right)$$

$$= 2 \sum_i \sum_j y_{ij} \log \left(\frac{y_{ij}}{\hat{\pi}_{ij}(H_0)\, y_{..}} \right) = 2 \sum_i \sum_j \text{ov} \log \left(\frac{\text{ov}}{\text{fv}} \right)$$, as in R code

Independence fitted values

So the null hypothesis of independence is very strongly rejected.

An alternative is Pearson's chi-square test from MAST90105.

```
# pearson's chisquared stat
sum((ov-fv)^2/fv)

## [1] 62.81231

summary(ov)

## Call: xtabs(formula = y ~ quality + particle)
## Number of cases in table: 450
## Number of factors: 2
## Test for independence of all factors:
##  Chisq = 62.81, df = 1, p-value = 2.274e-15
```

Via Logistic Fit

Each column of or is a sample from a multinomial distribution
Here factor has 2 levels → binomial model

```
# product multinomial model
(m <- matrix(y, nrow=2))
```

$= (or)^T$

```
##       [,1] [,2]
## [1,]  320   80
## [2,]   14   36
```

$[320 \;\; 80] \to y_1$

$[14 \;\; 36] \to y_2$

$y_j \sim \text{Multinomial}(y_{\cdot j}, P_j)$

$P_j = (P_{1j}, P_{2j})$; $j = 1, 2$

```
modb <- glm(m ~ 1, family=binomial)
deviance(modb)
```

– model H_0

```
## [1] 54.03045
```