Homework Set # 10 Solutions :

model : $\{n_{ij}\} \sim MULT(n, \{\pi_{ij}\})$, $H_0: \pi_{ij} = \pi_{i+}\pi_{+j}$

(X, Y are independent)

(1.) $X^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}}$, $\hat{m}_{ij} = \frac{n_{i+}n_{+j}}{n}$, $df = (I-1)(J-1)$

$X^2 = 16.955$, $df = 9$, p-value $= P(\chi^2_9 > 16.955) = .049$

The data provides moderate evidence against the independence model (in support of a general association model).

(2.) $\hat{\gamma} = .3604$, $\hat{\sigma}(\hat{\gamma}) = .1124$, $H_0: \gamma = 0$

$z^* = \frac{\hat{\gamma} - 0}{\hat{\sigma}(\hat{\gamma})} = 3.207$, p-value $= P(\chi^2_1 > (3.207)^2)$

$= .0014$

The data provides very strong evidence against the independence model (in support of a montonic (positive) association model).

(3.) Estimates from simpler models will have smaller variance than estimates from more complicated models. If the simpler model provides an appropriate fit to the data, then its estimates will also have small bias.

Bias can be thought of as model misspecification

Variance can be thought of as estimation error.

```
> obs = matrix(c(7,7,2,3,2,8,3,7,1,5,4,9,2,8,9,14),nrow = 4,byrow = TRUE)
> dimnames(obs) = list(husband=c("1","2","3","4"),
+                      wife=c("1","2","3","4"))
>
> print(obs)
       wife
husband 1 2 3  4
      1 7 7 2  3
      2 2 8 3  7
      3 1 5 4  9
      4 2 8 9 14

> I = dim(obs)[1]
> J = dim(obs)[2]
>
> total = sum(obs)
> row.sum = apply(obs,1,sum)
> col.sum = apply(obs,2,sum)
>
> expected = matrix(row.sum) %*% t(matrix(col.sum)) / total
> dimnames(expected) = dimnames(obs)
>
> print(expected,digits=6)
       wife
husband       1        2       3        4
      1 2.50549  5.84615 3.75824  6.89011
      2 2.63736  6.15385 3.95604  7.25275
      3 2.50549  5.84615 3.75824  6.89011
      4 4.35165 10.15385 6.52747 11.96703

> X2 = sum((obs-expected)^2/expected)
> df = (I-1)*(J-1)
> p.value.X2 = pchisq(X2,df,lower.tail = FALSE)
> print(X2)
[1] 16.95524
> print(df)
[1] 9
> print(p.value.X2)
[1] 0.0494215

> out = chisq.test(obs,correct = FALSE)

> out$observed
       wife
husband 1 2 3  4
      1 7 7 2  3
      2 2 8 3  7
      3 1 5 4  9
      4 2 8 9 14

> out$expected
       wife
husband        1         2        3         4
      1 2.505495  5.846154 3.758242  6.890110
      2 2.637363  6.153846 3.956044  7.252747
      3 2.505495  5.846154 3.758242  6.890110
      4 4.351648 10.153846 6.527473 11.967033

> out

        Pearson's Chi-squared test

data:  obs
X-squared = 16.955, df = 9, p-value = 0.04942
```

```
> library("MESS")

>
> gk.result = gkgamma(obs)
>
> gamma.hat = gk.result$estimate
> se = gk.result$se1
>
> z = gamma.hat / se
> p.value.g = 2*pnorm(abs(z),lower.tail = FALSE)
>
> g.table = matrix(c(gamma.hat,se,z,p.value.g),byrow=TRUE,nrow = 1)
> dimnames(g.table) =
+   list(c("Gamma"),c("estimate","ASE","z.stat","p.value"))

> print(g.table,digits = 4)
      estimate    ASE z.stat  p.value
Gamma   0.3604 0.1124  3.207 0.001341
```