Detergent Data (Table 5-1)

```
> # Navigate to the location of the data using the Misc menu
> soapdata <- read.table("DetergentFrame.txt"); soapdata
   Softness Prev_Use   Temp Pref Freq
1    1=Soft    1=Yes 1=High  1=X   19
2    1=Soft    1=Yes 1=High  2=M   29
3    1=Soft    1=Yes  2=Low  1=X   57
4    1=Soft    1=Yes  2=Low  2=M   49
5    1=Soft     2=No 1=High  1=X   29
6    1=Soft     2=No 1=High  2=M   27
7    1=Soft     2=No  2=Low  1=X   63
8    1=Soft     2=No  2=Low  2=M   53
9    2=Medm    1=Yes 1=High  1=X   23
10   2=Medm    1=Yes 1=High  2=M   47
11   2=Medm    1=Yes  2=Low  1=X   47
12   2=Medm    1=Yes  2=Low  2=M   55
13   2=Medm     2=No 1=High  1=X   33
14   2=Medm     2=No 1=High  2=M   23
15   2=Medm     2=No  2=Low  1=X   66
16   2=Medm     2=No  2=Low  2=M   50
17   3=Hard    1=Yes 1=High  1=X   24
18   3=Hard    1=Yes 1=High  2=M   43
19   3=Hard    1=Yes  2=Low  1=X   37
20   3=Hard    1=Yes  2=Low  2=M   52
21   3=Hard     2=No 1=High  1=X   42
22   3=Hard     2=No 1=High  2=M   30
23   3=Hard     2=No  2=Low  1=X   68
24   3=Hard     2=No  2=Low  2=M   42
> soap <- xtabs(Freq ~ Softness+Prev_Use+Temp+Pref, data=soapdata)
> summary(soap)
Call: xtabs(formula = Freq ~ Softness + Prev_Use + Temp + Pref, data = soapdata)
Number of cases in table: 1008
Number of factors: 4
Test for independence of all factors:
      Chisq = 43.9, df = 18, p-value = 0.0005957
> loglin(soap,list(1,2,3,4))$lrt # Matches text, p. 76
2 iterations: deviation 1.136868e-13
[1] 42.92866
```

```
> # Strategy: Find a model for the explanatory variables, using a
> # marginal table. Then check links of explanatory to response.
> soapex = xtabs(Freq ~ Softness+Prev_Use+Temp, data=soapdata); soapex
, , Temp = 1=High

        Prev_Use
Softness 1=Yes 2=No
  1=Soft    48   56
  2=Medm    70   56
  3=Hard    67   72

, , Temp = 2=Low

        Prev_Use
Softness 1=Yes 2=No
  1=Soft   106  116
  2=Medm   102  116
  3=Hard    89  110

> summary(soapex)
Call: xtabs(formula = Freq ~ Softness + Prev_Use + Temp, data = soapdata)
Number of cases in table: 1008
Number of factors: 3
Test for independence of all factors:
      Chisq = 10.019, df = 7, p-value = 0.1875
> soapexA = loglin(soapex,list(1,2,3)) # Complete independence
2 iterations: deviation 1.136868e-13
> soapexA$lrt
[1] 10.10304
>
> # Check 2-d marginal tables anyway
> softemp = xtabs(Freq ~ Softness+Temp, data=soapdata); softemp
        Temp
Softness 1=High 2=Low
  1=Soft    104   222
  2=Medm    126   218
  3=Hard    139   199
> round(100*prop.table(softemp,1),2) # Row percents
        Temp
Softness 1=High 2=Low
  1=Soft  31.90 68.10
  2=Medm  36.63 63.37
  3=Hard  41.12 58.88
> summary(softemp)
Call: xtabs(formula = Freq ~ Softness + Temp, data = soapdata)
Number of cases in table: 1008
Number of factors: 2
Test for independence of all factors:
      Chisq = 6.082, df = 2, p-value = 0.04778
> # Harder water goes with higher temp, sort of
```

```
> softprev = xtabs(Freq ~ Softness+Prev_Use, data=soapdata); softprev
        Prev_Use
Softness 1=Yes 2=No
  1=Soft   154  172
  2=Medm   172  172
  3=Hard   156  182
> round(100*prop.table(softprev,1),2) # Row percents
        Prev_Use
Softness 1=Yes  2=No
  1=Soft 47.24 52.76
  2=Medm 50.00 50.00
  3=Hard 46.15 53.85
> summary(softprev)
Call: xtabs(formula = Freq ~ Softness + Prev_Use, data = soapdata)
Number of cases in table: 1008
Number of factors: 2
Test for independence of all factors:
      Chisq = 1.0753, df = 2, p-value = 0.5841
> # Not much

> prevtemp = xtabs(Freq ~ Prev_Use+Temp, data=soapdata); prevtemp
        Temp
Prev_Use 1=High 2=Low
   1=Yes    185   297
   2=No     184   342
> summary(prevtemp)
Call: xtabs(formula = Freq ~ Prev_Use + Temp, data = soapdata)
Number of cases in table: 1008
Number of factors: 2
Test for independence of all factors:
      Chisq = 1.2535, df = 1, p-value = 0.2629
> # Not much
>
> JustSoftemp = loglin(soapex,list(2,c(1,3)))
2 iterations: deviation 0
> JustSoftemp$lrt; JustSoftemp$df
[1] 4.003931
[1] 5
> 1-pchisq(JustSoftemp$lrt, JustSoftemp$df)
[1] 0.5488501
> # Fits fine. Any better than complete independence?
> G2Change = soapexA$lrt-JustSoftemp$lrt; G2Change
[1] 6.099104
> dfChange = soapexA$df-JustSoftemp$df; dfChange
[1] 2
> pvalChange = 1-pchisq(G2Change, df=dfChange)
> pvalChange
[1] 0.04738014
> # Okay, keep [Softness Temperature]
>
```

```
> # Any IV, DV link at all?
> ModelA = loglin(soap,list(2,4,c(1,3))); ModelA
2 iterations: deviation 5.684342e-14
$lrt
[1] 36.82955

$pearson
[1] 37.76417

$df
[1] 16

$margin
$margin[[1]]
[1] "Prev_Use"

$margin[[2]]
[1] "Pref"

$margin[[3]]
[1] "Softness" "Temp"


> 1-pchisq(ModelA$lrt,ModelA$df)
[1] 0.002216038
> # Something is going on. Try model with all 2-way links
> # between explanatory and response variables.
> link2 = loglin(soap,list(c(1,3),c(1,4),c(2,4),c(3,4))); link2
3 iterations: deviation 0.06630545
$lrt
[1] 11.54287

$pearson
[1] 11.45839

$df
[1] 12

$margin
$margin[[1]]
[1] "Softness" "Temp"

$margin[[2]]
[1] "Softness" "Pref"

$margin[[3]]
[1] "Prev_Use" "Pref"

$margin[[4]]
[1] "Temp" "Pref"


> # Fits well. Try adding each link separately, and compare
```

```
> loglin(soap,list(c(1,3),c(1,4)))$lrt
2 iterations: deviation 1.421085e-14
[1] 38.3555
> loglin(soap,list(c(1,3),c(2,4)))$lrt
2 iterations: deviation 5.684342e-14
[1] 16.24809
> loglin(soap,list(c(1,3),c(3,4)))$lrt
2 iterations: deviation 0
[1] 34.3892
>

> ModelB = loglin(soap,list(c(1,3),c(2,4))) # [Soft Temp] [PrevUse Pref]
2 iterations: deviation 5.684342e-14
> # Does it fit?
> ModelB$lrt; ModelB$df
[1] 16.24809
[1] 15
> 1-pchisq(ModelB$lrt, ModelB$df)
[1] 0.365758
> # Improvement?
> G2Change = ModelA$lrt-ModelB$lrt; G2Change
[1] 20.58147
> dfChange = ModelA$df-ModelB$df; dfChange
[1] 1
> pvalChange = 1-pchisq(G2Change, df=dfChange); pvalChange
[1] 5.71467e-06
> # I like this one. But just check to see if another link is justified.
>
> loglin(soap,list(c(1,3),c(2,4),c(1,4)))$lrt # Add [Soft Pref]?
2 iterations: deviation 2.842171e-14
[1] 15.85279
> loglin(soap,list(c(1,3),c(2,4),c(3,4)))$lrt # Add [Temp Pref]?
2 iterations: deviation 5.684342e-14
[1] 11.88649
> ModelC = loglin(soap,list(c(1,3),c(2,4),c(3,4))) # Adding [Temp Pref]
2 iterations: deviation 5.684342e-14
> G2Change = ModelB$lrt-ModelC$lrt; G2Change
[1] 4.361601
> dfChange = ModelB$df-ModelC$df; dfChange
[1] 1
> pvalChange = 1-pchisq(G2Change, df=dfChange); pvalChange
[1] 0.03675775
> # I have to take it. Is link2 an improvement over this?
>
> ModelD = link2
> G2Change = ModelC$lrt-ModelD$lrt; G2Change
[1] 0.3436218
> dfChange = ModelC$df-ModelD$df; dfChange
[1] 2
> pvalChange = 1-pchisq(G2Change, df=dfChange); pvalChange
[1] 0.8421384
> # Okay, Model C looks like the choice.
> # [1 3] [2 4] [3 4] = [Soft Temp] [PrevUse Pref] [Temp Pref]
```

```
>
> # Look at marginal tables and parameter estimates to see what's happening
> PrevusePref = xtabs(Freq ~ Prev_Use+Pref, data=soapdata); PrevusePref
        Pref
Prev_Use 1=X 2=M
   1=Yes 207 275
   2=No  301 225
> round(100*prop.table(PrevusePref,1),2) # Row percents
        Pref
Prev_Use   1=X   2=M
   1=Yes 42.95 57.05
   2=No  57.22 42.78
> summary(PrevusePref)
Call: xtabs(formula = Freq ~ Prev_Use + Pref, data = soapdata)
Number of cases in table: 1008
Number of factors: 2
Test for independence of all factors:
      Chisq = 20.512, df = 1, p-value = 5.925e-06
> # Those who used M before tend to prefer it
> TempPref = xtabs(Freq ~ Temp+Pref, data=soapdata); TempPref
        Pref
Temp     1=X 2=M
  1=High 170 199
  2=Low  338 301
> round(100*prop.table(TempPref,1),2) # Row percents
        Pref
Temp       1=X   2=M
  1=High 46.07 53.93
  2=Low  52.90 47.10
> summary(TempPref)
Call: xtabs(formula = Freq ~ Temp + Pref, data = soapdata)
Number of cases in table: 1008
Number of factors: 2
Test for independence of all factors:
      Chisq = 4.358, df = 1, p-value = 0.03683
> # High temp goes with pref for M
```

```
> # Parameter estimates for Model C
> loglin(soap,list(c(1,3),c(2,4),c(3,4)),param=T)$param
2 iterations: deviation 5.684342e-14

$Softness.Temp
        Temp
Softness       1=High        2=Low
  1=Soft -0.101588153  0.101588153
  2=Medm  0.003448510 -0.003448510
  3=Hard  0.098139643 -0.098139643

$Prev_Use.Pref
        Pref
Prev_Use        1=X        2=M
   1=Yes -0.1437655  0.1437655
   2=No   0.1437655 -0.1437655

$Temp.Pref
        Pref
Temp            1=X        2=M
  1=High -0.0683605  0.0683605
  2=Low   0.0683605 -0.0683605

> #
> # Conclusions
> #
> #   Consumers with harder water tend to use higher temperature
> #   Those who used Brand M before tend to prefer it
> #   Use of High temperature water goes with preference for M
> #
> #   Book arrives at the same model
> #   But if the conclusions are actually stated in the book, I missed it.
```