```
> options(prompt = " ", continue = " ")
 data <- read.csv(file = "http://www.chrisbilder.com/categorical/Chapter3/pol_ideol_data.csv")
 str(data)

'data.frame':        20 obs. of  4 variables:
 $ gender: Factor w/ 2 levels "F","M": 1 1 1 1 1 1 1 1 1 1 ...
 $ party : Factor w/ 2 levels "D","R": 1 1 1 1 1 2 2 2 2 2 ...
 $ ideol : Factor w/ 5 levels "M","SC","SL",..: 5 3 1 2 4 5 3 1 2 4 ...
 $ count : int  44 47 118 23 32 18 28 86 39 48 ...


 mar <- xtabs(count ~ gender + party, data = data)
 odds_F = mar[1,1] / mar[1,2]
 odds_M = mar[2,1] / mar[2,2]
 OR = odds_F / odds_M
 tab = xtabs(count ~ gender + party + ideol, data = data)
 obs <- ftable(tab)
 m_OR = (obs[1,] / obs[2,]) / (obs[3,]/obs[4,])
```

PART A :
1. the sample odds ratios of Democratic vs Republican between females and males is 1.3819; it means that female has more chance to be Democratic than male.
2. The marginal ORs are 1.6051, 1.4744, 0.8887, 1.4783, and 0.8148 for political odeologies VL, SL, M, SC, and VC. For given political ideology VL, SL, and SC, the odds are greater than 1, which says that Females have more chance to be Democratic, but the opposite is true for political ideology M and VC.

```
 mantelhaen.test(tab)

        Mantel-Haenszel chi-squared test with continuity correction

data:  tab
Mantel-Haenszel X-squared = 3.2436, df = 1, p-value = 0.0717
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
 0.987835 1.754986
sample estimates:
common odds ratio
         1.316676
```

3. The test fail to reject the null hypthesis, which means female and political party are not independent given political ideology.

```
 ######################################################################
 # Function to perform the Breslow and Day (1980) test including
 # the corrected test by Tarone
 # Uses the equations in Lachin (2000) p. 124-125.
 #
 # Programmed by Michael Hoehle <http://www-m4.ma.tum.de/pers/hoehle>
 # Note that the results of the Tarone corrected test do
 # not correspond to the numbers in the Lachin book...
 #
```

```
# Params:
#  x - a 2x2xK contingency table
#
# Returns:
#  a vector with three values
#   1st value is the Breslow and Day test statistic
#   2nd value is the correct test by Tarone
#   3rd value - p value based on the Tarone test statistic
#               using a \chi^2(K-1) distribution
######################################################################

breslowday.test <- function(x) {
  #Find the common OR based on Mantel-Haenszel
  or.hat.mh <- mantelhaen.test(x)$estimate
  #Number of strata
  K <- dim(x)[3]
  #Value of the Statistic
  X2.HBD <- 0
  #Value of aj, tildeaj and Var.aj
  a <- tildea <- Var.a <- numeric(K)

  for (j in 1:K) {
    #Find marginals of table j
    mj <- apply(x[,,j], MARGIN=1, sum)
    nj <- apply(x[,,j], MARGIN=2, sum)

    #Solve for tilde(a)_j
    coef <- c(-mj[1]*nj[1] * or.hat.mh, nj[2]-mj[1]+or.hat.mh*(nj[1]+mj[1]),
                 1-or.hat.mh)
    sols <- Re(polyroot(coef))
    #Take the root, which fulfills 0 < tilde(a)_j <= min(n1_j, m1_j)
    tildeaj <- sols[(0 < sols) &  (sols <= min(nj[1],mj[1]))]
    #Observed value
    aj <- x[1,1,j]

    #Determine other expected cell entries
    tildebj <- mj[1] - tildeaj
    tildecj <- nj[1] - tildeaj
    tildedj <- mj[2] - tildecj

    #Compute \hat{\Var}(a_j | \widehat{\OR}_MH)
    Var.aj <- (1/tildeaj + 1/tildebj + 1/tildecj + 1/tildedj)^(-1)

    #Compute contribution
    X2.HBD <- X2.HBD + as.numeric((aj - tildeaj)^2 / Var.aj)

    #Assign found value for later computations
    a[j] <- aj ;  tildea[j] <- tildeaj ; Var.a[j] <- Var.aj
  }

  #Compute Tarone corrected test
```

```
  X2.HBDT <-as.numeric( X2.HBD -  (sum(a) - sum(tildea))^2/sum(Var.aj) )

  #Compute p-value based on the Tarone corrected test
  p <- 1-pchisq(X2.HBDT, df=K-1)

  res <- list(X2.HBD=X2.HBD,X2.HBDT=X2.HBDT,p=p)
  class(res) <- "bdtest"
  return(res)
}
print.bdtest <- function(x) {
  cat("Breslow and Day test (with Tarone correction):\n")
  cat("Breslow-Day X-squared         =",x$X2.HBD,"\n")
  cat("Breslow-Day-Tarone X-squared  =",x$X2.HBDT,"\n\n")
  cat("Test for test of a common OR: p-value = ",x$p,"\n\n")
}


breslowday.test(tab)
```

```
Breslow and Day test (with Tarone correction):
Breslow-Day X-squared         = 3.235357
Breslow-Day-Tarone X-squared  = 3.23528

Test for test of a common OR: p-value =  0.5192516
```

4. From the output above, the test fail to reject the null hypothesis.

$$\chi^2 = \sum_{k=1}^{5} \frac{(m_{11k} - A_k)^2}{V_k}$$
$$A_k = \frac{m_{1+k} m_{+1k}}{m_{++k}}$$
$$V_k = (1/A_k + 1/B_k + 1/C_k + 1/D_k)^{-1}$$
$$B_k = m_{1+k} - A_i$$
$$C_k = m_{+1k} - A_k$$
$$D_k = m_{++k} - A_k - B_k - C_k$$

PART B

14 (a)
Use factor to make VL < SL < M < SC < VC

```
v <- as.ordered(c("VL", "SL", "M", "SC", "VC"))
data[,3] = factor(data[,3], v, ordered = TRUE)
```

14 (b)

```
obs <- ftable(xtabs(count ~ gender + party + ideol, data = data))
obs
```

```
              ideol  VL  SL   M  SC  VC
gender party
F      D             44  47 118  23  32
       R             18  28  86  39  48
M      D             36  34  53  18  23
       R             12  18  62  45  51
```

14 (c) i.

```
library(nnet)
library(car)
library(package = MASS)
mod.m <- multinom(formula = ideol ~ gender + party + gender*party,
                  weights = count, data = data)
```

```
# weights:  25 (16 variable)
initial  value 1343.880657
iter  10 value 1231.244704
iter  20 value 1229.548447
final  value 1229.543342
converged
```

```
mod.ph <- polr(formula = ideol ~ gender + party + gender*party,
               weights = count, data = data, method = 'logistic')
lr1 <- Anova(mod.m)
lr2 <- Anova(mod.ph)
c = round(coefficients(mod.m),4)
c2 = round(coefficients(mod.ph),4)
s = summary(mod.ph)
intercpt <- s$coefficients
```

The coefficents for the two models are

```
c
```

```
   (Intercept) genderM partyR genderM:partyR
SL      0.0660 -0.1232 0.3759         0.0868
M       0.9865 -0.5998 0.5775         0.6780
SC     -0.6487 -0.0444 1.4219         0.5929
VC     -0.3184 -0.1297 1.2992         0.5958
```

```
for (i in 1:4){
print(sprintf("log(%8s) = %6.3f + %6.3fG + %6.3fP + %6.3fG:P",
              paste("p",v[i+1],"/", "p",v[i], sep=""), c[i,1], c[i,2], c[i,3], c[i,4]))
}
```

```
[1] "log( pSL/pVL) =  0.066 + -0.123G +  0.376P +  0.087G:P"
[1] "log(  pM/pSL) =  0.987 + -0.600G +  0.578P +  0.678G:P"
[1] "log(  pSC/pM) = -0.649 + -0.044G +  1.422P +  0.593G:P"
[1] "log( pVC/pSC) = -0.318 + -0.130G +  1.299P +  0.596G:P"
```

```
c2
```

```
      genderM         partyR genderM:partyR
      -0.1431         0.7562         0.5091
```

```
for (i in 4:7){
print(sprintf("log(%8s) = %6.3f + %6.3fg + %6.3fP + %6.3fG:P",
              paste("p",v[i+1],"/", "p",v[i], sep=""), -intercpt[i,2], -c2[1], -c2[2], -c2[3]))
}
```

```
[1] "log( pVC/pSC) = -0.133 +  0.143g + -0.756P + -0.509G:P"
[1] "log( pNA/pVC) = -0.116 +  0.143g + -0.756P + -0.509G:P"
[1] "log( pNA/pNA) = -0.123 +  0.143g + -0.756P + -0.509G:P"
[1] "log( pNA/pNA) = -0.136 +  0.143g + -0.756P + -0.509G:P"
```

LR test for the two models

```
lr1
```

```
Analysis of Deviance Table (Type II tests)

Response: ideol
             LR Chisq Df Pr(>Chisq)
gender          8.965  4    0.06198 .
party          60.555  4  2.218e-12 ***
gender:party    3.245  4    0.51763
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

```
lr2
```

```
Analysis of Deviance Table (Type II tests)

Response: ideol
             LR Chisq Df Pr(>Chisq)
gender          0.843  1    0.35864
party          56.847  1  4.711e-14 ***
gender:party    3.992  1    0.04571 *
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

14 (c) iii.

```
P1 = unique(mod.m$fitted.values)
P2 = unique(mod.ph$fitted.values)
e1 = c(t(mar)) * P1
e2 = c(t(mar)) * P2
```

Marginal counts from model multi.

```
row.names(e1) <- c("F D", "F R", "M D", "M R")
round(e1,0)
```

```
    VL SL   M SC VC
F D 44 47 118 23 32
F R 18 28  86 39 48
M D 36 34  53 18 23
M R 12 18  62 45 51
```

Marginal counts from prop. odds. reg.

```
row.names(e2) <- c("F D", "F R", "M D", "M R")
round(e2,0)
```

```
    VL SL   M SC VC
F D 46 50 105 31 31
F R 20 27  85 38 49
M D 32 33  64 18 17
M R 12 17  66 37 55
```

The observed counts

```
obs
```

```
             ideol  VL  SL   M  SC  VC
gender party
F      D            44  47 118  23  32
       R            18  28  86  39  48
M      D            36  34  53  18  23
       R            12  18  62  45  51
```

14 (c) iv
From the obove, we see that obs and multi. model have the same counts, because the mult. model is a saturated model, meaning that it has more than enought parameters.

14 (d)
Since the estimated counts from the two models are close, so the two models similar.

15
The proportional odds model examines the odds of a response with respect to cumulative probabilities. To understand the equivalent with respect to the model estimated counts in a three-way contingency table structure, define m_{ijk} as the estimated count for gender i ($i = 1$ corresponds to female and $i = 2$ corresponds to male), party j ($j = 1$ corresponds to Democrat and $j = 2$ corresponds to Republican), and ideology level k ($k = 1$ corresponds to very liberal,..., and $k = 5$ corresponds to very conservative). For female Democrats, the estimated odds of a very liberal response are

$(m111/m11+)/(m112 + m113 + m114 + m115)/m11+) = 46.14/(50.14 + 104.93 + 31.35 + 31.44) = 0.2118.$

From the model, the estimated odds are

$$OddsFD(Y1) = exp(\beta_{10} + \beta_1 R + \beta_2 M + \beta_3 RM).$$

For female Democrats, the two indicator variables have values of 0, so the estimated odds are $\exp(\beta_{10}) = \exp(1.5521) = 0.2118$. As an additional example for female Democrats, the estimated odds of a slightly liberal or lower response can be calculated as

$((m111+m112)/m11+)/(m113+m114+m115)/m11+) = (46.14+50.14)/(104.93+31.35+31.44) = 0.5741$

and $\exp(\beta_{20}) = \exp(0.5550) = 0.5741$. Similar calculations can be performed for other odds.