Homework #6

2. Refer to the Pregnancy Duration Data (p. 609), repeat the analysis on p.613 (the response variable is treated as Nominal categorical) using R or other statistical software. Compare your results with the ones in the text (from Minitab). Are they the same? If not, what is the cause? Interpret the parameters in the context of the problem.

```
reg2 <- multinom(cbind(preg3,preg2,preg1) ~ .-preg, data = df2)

## # weights:  21 (12 variable)
## initial  value 112.058453
## iter  10 value 84.619847
## final  value 84.337718
## converged

summary(reg2)

## Call:
## multinom(formula = cbind(preg3, preg2, preg1) ~ . - preg, data = df2)
##
## Coefficients:
##       (Intercept)       nutri     age1     age3  alcohol  smoking
## preg2    3.958370 -0.04644903 2.913475 1.887550 1.067001 2.230492
## preg1    5.475147 -0.06541919 2.957028 2.059662 2.042900 2.452362
##
## Std. Errors:
##       (Intercept)      nutri      age1      age3   alcohol   smoking
## preg2    1.941063 0.01488581 0.8575544 0.8088255 0.6495262 0.6681955
## preg1    2.271677 0.01823916 0.9644921 0.8947727 0.7097461 0.7315106
##
## Residual Deviance: 168.6754
## AIC: 192.6754
```

The results are the same. Actually I set the 3^{rd} category of `preg` as the reference category to make them the same. After adjusting other factors, when nutrition status increases 1 unit, the odds of pregnancy duration in category 2 over category 3 will change by a factor of $e^{-0.046}$, and the odds of pregnancy duration in category 1 over category 3 will change by a factor of $e^{-0.065}$. Keeping other variables constant, when a mother's age change from age category 2 to age category 1, the odds of pregnancy duration in category 2 over category 3 will change by a factor of $e^{2.91}$. Similarly, interpretation for other parameters can be drawn according to the output.

3. (8.2) The data in Table 8.5 are from an investigation into satisfaction with hous- ing conditions in Copenhagen (derived from Example W in Cox and Snell, 1981, from original data from Madsen, 1971). Residents in selected areas living in rented homes built

between 1960 and 1968 were questioned about their satisfaction and the degree of contact with other residents. The data were tabulated by type of housing.
In addition, test whether there is interaction effect between "type of housing" and "contact with other neighbors" on the response variable "satisfaction."

Table 8.5 *Satisfaction with housing conditions.*

Contact with other residents	Satisfaction: Low		Satisfaction: Medium		Satisfaction: High	
	Low	High	Low	High	Low	High
Tower block	65	34	54	47	100	100
Apartment	130	141	76	116	111	191
House	67	130	48	105	62	104

a. Summarize the data using appropriate tables of percentages to show the associations between levels of satisfaction and contact with other residents, levels of satisfaction and type of housing, and contact and type of housing.

```
##             contact
## satisfaction      high       low
##       high   0.2349792 0.1624033
##       low    0.1814396 0.1558596
##       medium 0.1594289 0.1058894
##             type
## satisfaction  Apartment      House TowerBlock
##       high   0.17965497 0.09875074 0.11897680
##       low    0.16121356 0.11719215 0.05889352
##       medium 0.11421773 0.09101725 0.06008328
##        type
## contact Apartment     House TowerBlock
##    high 0.2665080 0.2016657  0.1076740
##    low  0.1885782 0.1052945  0.1302796
```

b. Use nominal logistic regression to model associations between level of satisfaction and the other two variables. Obtain a parsimonious model that summarizes the patterns in the data.

```
reg3 <- multinom(satisfaction ~ contact + type, weights = frequency, data = df3)

## # weights:  15 (8 variable)
## initial  value 1846.767257
## iter  10 value 1803.151908
## final  value 1802.740161
## converged

summary(reg3)
```

```
## Call:
## multinom(formula = satisfaction ~ contact + type, data = df3,
##     weights = frequency)
##
## Coefficients:
##        (Intercept) contactlow typeHouse typeTowerBlock
## low     -0.2474055  0.3282260 0.3040225     -0.6415725
## medium  -0.4654412  0.0322483 0.3736997     -0.2348298
##
## Std. Errors:
##        (Intercept) contactlow typeHouse typeTowerBlock
## low     0.09783067  0.1181870 0.1351693       0.150077
## medium  0.10466300  0.1269192 0.1454812       0.154099
##
## Residual Deviance: 3605.48
## AIC: 3621.48
```

```
null <- multinom(satisfaction ~ 1, weights = frequency, data = df3)
```

```
## # weights:  6 (2 variable)
## initial  value 1846.767257
## final  value 1824.438811
## converged
```

```
summary(null)
```

```
## Call:
## multinom(formula = satisfaction ~ 1, data = df3, weights = frequency)
##
## Coefficients:
##        (Intercept)
## low     -0.1639289
## medium  -0.4039694
##
## Std. Errors:
##        (Intercept)
## low     0.05710231
## medium  0.06114866
##
## Residual Deviance: 3648.878
## AIC: 3652.878
```

* Test whether there is interaction effect between "type of housing" and "contact with other neighbors" on the response variable "satisfaction".

```r
full <-  multinom(satisfaction ~ contact*type, weights = frequency, data = df3)
```

```
## # weights:  21 (12 variable)
## initial  value 1846.767257
## iter  10 value 1800.889217
## final  value 1799.293647
## converged
```

```r
c('Chi^2' = reg3$dev-full$dev,'p-value' = 1-pchisq(reg3$dev-full$dev, 10))
```

```
##     Chi^2   p-value
## 6.8930278 0.7355037
```

```r
# Df=12-2
```

The likelihood ratio test indicates no interaction effect between "type of housing" and "contact with other neighbors" on the response variable "satisfaction".