מודלים סטטיסטיים – מבחן מועד א' תשע"ב – פתרון מקוצר (ראשי פרקים)

שאלה 1

```
Dependent Variable: y

                                                Sum of
Source                          DF             Squares     Mean Square    F Value    Pr > F

Model                            A                   C               D          F    XXXXXX

Error                            B     11.33252000                   E

Corrected Total                 XX     XXXXXXXXXXX      XXXXXXXXXX


                                                      Standard
Parameter                         Estimate               Error    t Value    Pr > |t|

beethoven                                G                   H          I    XXXXXXXX


         Levene's Test XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
        XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

                              Sum of         Mean
Source          DF           Squares       Square    F Value    Pr > F

group           XX                 J       XXXXXX    XXXXXXX    XXXXXX
Error           XX            2.6300       XXXXXX
```

$$\text{A} = I - 1 = 3 - 1 = 2$$

$$\text{B} = N - I = 15 - 3 = 12$$

$$\text{C} = \sum_{i=1}^{I} n_i (\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2 = 14.3507$$

$$\text{D} = \text{C} / \text{A} = 14.3507 \; / \; 2 = 7.1753$$

$$\text{E} = 11.3325 / 12 = 0.9444$$

$$\text{F} = \text{D} / \text{E} = 7.60$$

$$\text{G} = (-0.5)\overline{Y}_{1\cdot} + (-0.5)\overline{Y}_{2\cdot} + \overline{Y}_{3\cdot} = (-0.5)(1.726) + (-0.5)(0.01) + (-0.58) = -1.448$$

$$\text{H} = s\left[\sum_{i=1}^{I} c_i^2 / n_i\right]^{1/2} = \sqrt{0.9444}\,[(-0.5)^2/5 + (-0.5)^2/5 + 1/5]^{1/2} = 0.5323$$

$$\text{I} = \text{G} / \text{H} = -2.72$$

$$J = \sum_{i=1}^{I} n_i (\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})^2 = 0.7876, \quad X_{ij} = |Y_{ij} - \overline{Y}_{i\cdot}|$$

<u>שאלה 2</u>

המודל (AB,AC,D) אומר

$$\log \pi_{ijkl} = \overline{\theta}_{\cdots\cdot} + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{ij}^{AB} + \lambda_{ik}^{AC}$$

הפירוש הוא ש- (א) B ו- C ב״ת בהינתן A ו- (ב) D ב״ת בשלשת המשתנים A, B, C. דהיינו

$$\Pr(A=i,\ B=j,\ C=k,\ D=l) \ = \ \Pr(A=i)\Pr(B=j \mid A=i)\Pr(C=j \mid A=i)\Pr(D=l)$$

דהיינו

$$\pi_{ijkl} = \frac{\pi_{ij\cdot\cdot}\pi_{i\cdot k\cdot}\pi_{\cdot\cdot\cdot l}}{\pi_{i\cdot\cdot\cdot}}$$

ההוכחה דומה להוכחות שנעשו בשיעורים ותרגילים עבור מודלים דומים.

שאלה 3

מדובר בניתוח שונות דו-כיווני לא מאוזן והפתרון הינו באמצעות מטריצות כמו בשאלה 2 של בוחן 2.
החישובים להלן (בוצעו בתוכנה R):

```
> x0 = rep(1,14)
> x1 = c(rep(1,7),rep(-1,7))
> x2 = c(rep(1,3),rep(-1,4),rep(1,4),rep(-1,3))
> x = cbind(x0,x1,x2)
> x
      x0 x1 x2
 [1,]  1  1  1
 [2,]  1  1  1
 [3,]  1  1  1
 [4,]  1  1 -1
 [5,]  1  1 -1
 [6,]  1  1 -1
 [7,]  1  1 -1
 [8,]  1 -1  1
 [9,]  1 -1  1
[10,]  1 -1  1
[11,]  1 -1  1
[12,]  1 -1 -1
[13,]  1 -1 -1
[14,]  1 -1 -1
> y = c(23,26,29,11,12,14,17,33,36,38,41,22,24,27)
> dim(y) = c(14,1)
> c = t(x) %*% x
> c
   x0 x1 x2
x0 14  0  0
x1  0 14 -2
x2  0 -2 14
> ci = solve(c)
> ci
           x0         x1         x2
x0 0.07142857 0.00000000 0.00000000
x1 0.00000000 0.07291667 0.01041667
x2 0.00000000 0.01041667 0.07291667
> h = t(x) %*% y
> h
   [,1]
x0  353
x1  -89
x2   99
> etahat = ci %*% h
> etahat
        [,1]
x0 25.214286
x1 -5.458333
x2  6.291667
> yhat = x %*% etahat
> yhat
          [,1]
 [1,] 26.04762
 [2,] 26.04762
 [3,] 26.04762
 [4,] 13.46429
 [5,] 13.46429
 [6,] 13.46429
 [7,] 13.46429
```

```
 [8,] 36.96429
 [9,] 36.96429
[10,] 36.96429
[11,] 36.96429
[12,] 24.38095
[13,] 24.38095
[14,] 24.38095
> y1b = sum(y[1:3,]) + sum(y[8:11,])
> y1b = y1b/7
> y2b = sum(y[4:7,]) + sum(y[12:14,])
> y2b = y2b/7
> print(cbind(y1b,y2b))
          y1b      y2b
[1,] 32.28571 18.14286
> yhat0 = c(rep(y1b,3),rep(y2b,4),rep(y1b,4),rep(y2b,3))
> dim(yhat0) = c(14,1)
> yhat0
          [,1]
 [1,] 32.28571
 [2,] 32.28571
 [3,] 32.28571
 [4,] 18.14286
 [5,] 18.14286
 [6,] 18.14286
 [7,] 18.14286
 [8,] 32.28571
 [9,] 32.28571
[10,] 32.28571
[11,] 32.28571
[12,] 18.14286
[13,] 18.14286
[14,] 18.14286
> diff = yhat-yhat0
> diff
           [,1]
 [1,] -6.238095
 [2,] -6.238095
 [3,] -6.238095
 [4,] -4.678571
 [5,] -4.678571
 [6,] -4.678571
 [7,] -4.678571
 [8,]  4.678571
 [9,]  4.678571
[10,]  4.678571
[11,]  4.678571
[12,]  6.238095
[13,]  6.238095
[14,]  6.238095
> ssa = sum(diff^2)
> ssa
[1] 408.5952
> e = y-yhat
> e
             [,1]
 [1,] -3.04761905
 [2,] -0.04761905
 [3,]  2.95238095
 [4,] -2.46428571
 [5,] -1.46428571
 [6,]  0.53571429
 [7,]  3.53571429
```

```
 [8,] -3.96428571
 [9,] -0.96428571
[10,]  1.03571429
[11,]  4.03571429
[12,] -2.38095238
[13,] -0.38095238
[14,]  2.61904762
> sse = sum(e^2)
> sse
[1] 85.69048
> msa = ssa
> mse = sse/(14-3)
> mse
[1] 7.790043
> f = msa/mse
> f
[1] 52.45096
```

<u>שאלה 4</u>

פונקציית הנראות הינה (כאן שמים $X_{i0}=1$)

$$L(\beta)=\prod_{i=1}^{N} f_{Y_i}(Y_i)=\prod_{i=1}^{N}\left[\binom{n_i}{Y_i} p(X_i,\beta)^{Y_i}(1-p(X_i,\beta))^{n_i-Y_i}\right]=\prod_{i=1}^{N}\left[\binom{n_i}{Y_i} e^{Y_i\beta^T X_i}(1+e^{Y_i\beta^T X_i})^{-n_i}\right]$$

יש לנו

$$\ell(\beta)=\log L(\beta)=\sum_{i=1}^{N}\left[\log\binom{n_i}{Y_i}+Y_i\beta^T X_i-n_i\log(1+e^{Y_i\beta^T X_i})\right]$$

$$\frac{\partial \ell}{\partial \beta_r}=\sum_{i=1}^{n} X_{ir}\left[Y_i-n_i\frac{e^{Y_i\beta^T X_i}}{1+e^{Y_i\beta^T X_i}}\right]$$

צריכים לפתור את המשוואות

$$\frac{\partial \ell}{\partial \beta_r}=0, \quad r=0,\dots,p$$

עושים זאת באמצעות נוטון-רפסון באופן כמעט (אמנם לא לגמרי) זהה למה שעשינו בכיתה עבור רגרסיה לוגיסטית.