Northwest Atlantic   Fisheries Organization

SCIENTIFIC COUNCIL MEETING – NOVEMBER 2003

A Short Note on Modelling in S-plus the Standardized CPUE for
Northern Shrimp (*Pandalus borealis)* in Division 3M

by

A. Gudmundsdottir

Marine Research Institute, Reykjavik, Iceland

Abstract

The standardized catch per unit effort (CPUE) model for Northern Shrimp (*Pandalus borealis)* in Division 3M has been run in SAS until 2002. CPUE was modelled against year, month, vessel, area and gear. Now the standardized CPUE is modelled in S-plus. The results were the same using both softwares so Splus could be used in the next years.

Introduction

Up until 2002 the standardized catch per unit effort (CPUE) model for Northern Shrimp (*Pandalus borealis)* in Division 3M has been run in SAS, by using the Proc GENMOD (U. Skuladottir and D. Orr, 2002) module. CPUE was modelled against year, month, vessel, area and gear. Now the intention is to model the standardized CPUE in S-plus.

Material and Methods

A standardized datafile *nations02.csv*, the SAS program for making the CPUE-model and the results file was made available by D. Orr. The standardized datafile consists of data from 1993 to 2002 from Canada, Faroe Islands, Greenland, Iceland, Norway and Russia. The data are aggregated data of catch in kilos, effort in hours and number of tows by vessel, horsepower, tonnage, year, month, area, gear and nation. The data in the standardized datafile was imported in S-plus. Data were extracted from the datafile where catch >0 kg, effort >10 hours, gear was defined 1, 2 or 3 (single, double and triple). Area is allowed to have missing values. Horsepower and tonnage are not used in the model and many of the records have no values for these two variables.

The model should be standardized to the year 1993, the month June, single trawl and Icelandic data. In SAS the variables are standardized to the one with the highest numbers, but in S-plus it is the other way round, i.e. they are standardized to the lowest one, so as in SAS some recoding had to be done in S-plus. All factors have the same order in S-plus as in SAS. In S-plus (version 6) the GLM procedure was used, with the family parameter set to Gamma, the link was set to log and effort was used as a weighting factor.

Results

It was noted that in the standardized datafile, that the Russian catch was in tonnes as the other nations catches were in kilos. It was taken into account in the SAS program. Now the standardized datafile has been changed, so that all catches are in kilos.

A summary table (Table 1) was made with the extracted data and compared with the corresponding table from SAS. The results were the same for all years, except 4 and for those years the difference was so small, that it was concluded that the same data was being used in S-plus as in SAS. In 2002 CPUE was modelled against year, vessel, month, area and gear. As area is not always defined, data were extracted with the same condition as before, but area had to been defined too. A summary table (Table 2) was made for these data. The model was then fitted. The results from the model fit are shown in Table 3 and Fig. 1. The values of the estimate of the year factors are up to the 2nd and 3rd decimal the same as in SAS and therefore it is concluded that the model in Splus is giving the same result as the one in SAS.

The partial t-test provided little information about the area (Table 3). By using the *ANOVA* function in S-plus an analysis of deviance for the sequential addition of each variable was made (Table 4). It looks like that the variable *area* is not so important to the regression. Therefore another fit was made without the area and the two fits, that is with and without the area, compared by using ANOVA. The results are shown in Table 5. The conclusion drawn from this exercise is that there is no difference between these two models.

References

Skúladóttir, U. and D. Orr, 2002. The Assessment of the International Fishery for Shrimp (*Pandalus borealis*) in Division 3M (Flemish Cap), 1993-2002. *NAFO SCR Doc.,* 02/163, Serial No. N4793.

Statistical models in S. Chambers, J.M., and Hastie, T.J. 1992. Wadsworth & Brooks/Cole Advanced Books & Software. Pacific Grove, California.

S-PLUS 4 Guide to Statistics, Data Analysis Products Division, MathSoft, Seattle.

Table 1.   Analysis about the CPUE data. Numbers in italic are from the assessment in 2002, which differ from the one made in S-plus.

year	Number of obs.	Mean cpue	Std. dev.	Minimum	Maximum
1993	244	356.7318	147.12933	44.18182	894.5000
1994	230	236.9328	104.68254	10.37500	720.8765
1995	473 *471*	269.6544 *270.0095*	129.69249	38.14345	1181.9231
1996	927	228.1609 *227.9949*	114.52436	45.23596	847.5866
1997	378	285.7067	98.75083	44.30877	602.2971
1998	325 *324*	376.3973 *376.5447*	145.79688	34.38776	1315.7314
1999	338	382.7407	150.45038	35.25763	851.3818
2000	345	434.3164	173.21754	46.53540	1185.8592
2001	406	375.3852	176.72791	39.24800	1007.3333
2002	109 *110*	478.1836 *475.8319*	174.65671	123.88060	944.8406

Table 2. Analysis about the CPUE data, when area is defined.

year	Number of obs.	Mean cpue	Std. dev.	Minimum	Maximum
1993	170	386.2746	147.7421	92.81818	894.5000
1994	130	246.4198	126.2234	10.37500	720.8765
1995	362	276.7768	141.3902	38.14345	1181.9231
1996	863	229.4332	116.9914	45.23596	847.5866
1997	365	284.8268	99.6234	44.30877	602.2971
1998	316	377.0577	147.2426	34.38776	1315.7314
1999	328	381.2073	152.0707	35.25763	851.3818
2000	345	434.3164	173.2175	46.53540	1185.8592
2001	406	375.3852	176.7279	39.24800	1007.3333
2002	109	478.1836	174.6567	123.88060	944.8406

Table 3. Results from the multiplicative model. The ship factors are not shown.

```
Call: glm(formula = cpue ~ factor(year) + factor(ship.nr) + factor(month.nr) +
factor(area.nr) + factor(gear), family = Gamma(link = log), data = reg.data, weights =
effort)

Deviance Residuals:
      Min       1Q    Median       3Q       Max
 -31.55938 -1.778845 -0.257723 1.461865 19.98631

Coefficients:
                         Value              Std Error           t value
(Intercept)              6.5050031          0.06097707          106.679501
factor(year)1994         -0.5330655         0.05134741          -10.381547
factor(year)1995         -0.4069230         0.05140219          -7.916453
factor(year)1996         -0.5218728         0.05141306          -10.150587
factor(year)1997         -0.5398354         0.05298142          -10.189145
factor(year)1998         -0.2847182         0.05426353          -5.246953
factor(year)1999         -0.2685623         0.05434168          -4.942106
factor(year)2000         -0.1183063         0.05454599          -2.168927
factor(year)2001         -0.2147293         0.05822581          -3.687872
factor(year)2002         -0.1248797         0.06929505          -1.802145
factor(month.nr)2        -0.132094175       0.06678741          -1.97783034
factor(month.nr)3        0.043002208        0.05275844           0.81507737
factor(month.nr)4        -0.012386506       0.03074660          -0.40285770
factor(month.nr)5        -0.112538218       0.02579673          -4.36249986
factor(month.nr)6        -0.056981497       0.02301206          -2.47615768
factor(month.nr)7        -0.095276243       0.02146808          -4.43804129
factor(month.nr)8        -0.174796905       0.02522219          -6.93028282
factor(month.nr)9        -0.231001331       0.02649730          -8.71791939
factor(month.nr)10       -0.270494154       0.02878999          -9.39542287
factor(month.nr)11       -0.257347632       0.03225999          -7.97730115
factor(month.nr)12       -0.068453067       0.04704010          -1.45520664
factor(area.nr)2         -0.006135149       0.02154715          -0.28473132
factor(area.nr)3         -0.036099330       0.01759739          -2.05140309
factor(area.nr)4         -0.03065659        0.01716469          -1.786027
factor(gear)2            0.15291489         0.02946011           5.190574
factor(gear)3            -0.48625077        0.14735360          -3.299891

(Dispersion Parameter for Gamma family taken to be 16.79377)

    Null Deviance: 139844.1 on 3393 degrees of freedom

Residual Deviance: 57238.55 on 3230 degrees of freedom

Number of Fisher Scoring Iterations: 5
```

Table 4. Results from analysis of deviance.

```
Analysis of Deviance Table

Gamma model

Response: cpue

Terms added sequentially (first to last)
                  Df Deviance Resid. Df Resid. Dev  F Value      Pr(F)
          NULL                    3393   139844.1
    factor(year)   9 23647.65     3384   116196.5 156.4578 0.00000000
 factor(ship.nr) 138 54962.82     3246    61233.6  23.7160 0.00000000
factor(month.nr)  11  3096.89     3235    58136.8  16.7643 0.00000000
  factor(area.nr)   3   121.88     3232    58014.9   2.4192 0.06432497
     factor(gear)   2   776.31     3230    57238.6  23.1132 0.00000000
```

Table 5. Results comparing two models with anova.

Full model: factor(year) + factor(ship.nr) + factor(month.nr) + factor(area.nr) + factor(gear)
Reduced model: factor(year) + factor(ship.nr) + factor(month.nr) + factor(gear)

```
Analysis of Deviance Table

Response: cpue

Terms        Resid. Df Resid. Dev              Test Df  Deviance  F Value      Pr(F)
1 Full model    3230    57238.55
2 Reduced model 3233    57345.45  -factor(area.nr) -3 -106.8997 2.121813 0.09534365
```

Table 6. Results from the multiplicative model, when area is not used. The ship factors are not shown.

```
Call: glm(formula = cpue ~ factor(year) + factor(ship.nr) + factor(month.nr) +
factor(gear), family = Gamma(link = log), data = reg.data, weights = effort)
Deviance Residuals:
      Min        1Q   Median       3Q      Max
 -31.73736 -1.785077 -0.256367 1.460476 19.64501

Coefficients:
                      Value           Std.Error         t.value
(Intercept)           6.4866853       0.06057105        107.092180
factor(year)1994      -0.5336043      0.05119524        -10.422928
factor(year)1995      -0.4082135      0.05133204        -7.952411
factor(year)1996      -0.5222215      0.05145682        -10.148732
factor(year)1997      -0.5387834      0.05298747        -10.168128
factor(year)1998      -0.2812801      0.05429741        -5.180359
factor(year)1999      -0.2638335      0.05432468        -4.856605
factor(year)2000      -0.1172062      0.05460839        -2.146303
factor(year)2001      -0.2139889      0.05828739        -3.671273
factor(year)2002      -0.1226784      0.06936870        -1.768498
factor(month.nr)2     -0.131009859    0.06685524        -1.9596051
factor(month.nr)3     0.038226696     0.05265463        0.7259892
factor(month.nr)4     -0.009177865    0.03060537        -0.2998776
factor(month.nr)5     -0.109651642    0.02575334        -4.2577644
factor(month.nr)6     -0.055620294    0.02302179        -2.4159849
factor(month.nr)7     -0.095021533    0.02143382        -4.4332521
factor(month.nr)8     -0.175581348    0.02503975        -7.0121055
factor(month.nr)9     -0.231079460    0.02627433        -8.7948752
factor(month.nr)10    -0.270680343    0.02873377        -9.4202865
factor(month.nr)11    -0.263081649    0.03218004        -8.1753063
factor(month.nr)12    -0.069664807    0.04707090        -1.4799973
factor(gear)2         0.154707132     0.02949290        5.2455722
factor(gear)3         -0.488079786    0.14751771        -3.3086182

(Dispersion Parameter for Gamma family taken to be 16.83906)

    Null Deviance: 139844.1 on 3393 degrees of freedom

Residual Deviance: 57345.45 on 3233 degrees of freedom

Number of Fisher Scoring Iterations: 5
```
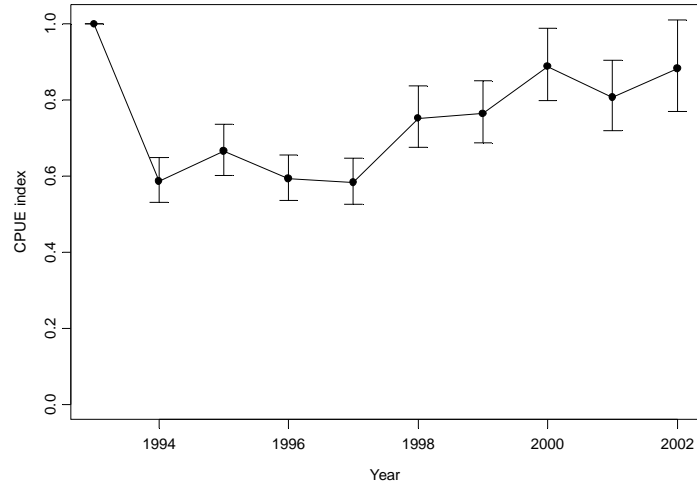
Fig. 1. Modelled CPUE index by year. Upper and lower bounds represent approximate 95 percent confidence limits.