"Where have all the puffins gone…"[1]

An Analysis of Atlantic Puffin (*Fratercula arctica)* Banding Returns in Newfoundland

Dave Fifield

[1] With apologies to Pete Seeger.

Introduction

Seabirds are the most visible inhabitants of the marine environment and many species have been extensively studied at their breeding colonies, but less is known of their ecology during the non-breeding season (Croxall 1984; Harrison 1983). This is certainly the case for the Atlantic alcidae and the Atlantic Puffin (*Fratercula arctica*) in particular (Lock et al. 1994; Brown 1986; Brown 1985; Nettleship and Evans 1985).

The Atlantic Puffin is the charismatic provincial bird of Newfoundland and Labrador and the focus of a thriving tourist industry. Its comical appearance and universal public appeal make it a powerful conservation icon. The North American breeding population is estimated at 350,000-400,000 pairs, with more than 90% (320,000-390,000) pairs breeding in Newfoundland and Labrador (Lowther et al. 2002).

The purpose of this study is to investigate movement and distribution patterns of Newfoundland and Labrador Atlantic Puffins through an analysis of banding recovery data. Bird band recoveries provide detailed movement data for many bird species (Dennis 1981; Mead 1974) and in North America, bird-banding records are jointly managed by the U.S. Fish and Wildlife Service and the Canadian Wildlife Service. The most recent analysis of North American Puffin band returns is more than 20 years old (Harris 1984).

Many aspects of the physical environment and of puffin behavior likely act in concert to influence band-return patterns. This paper addresses the extent to which relative frequency of band-returns and distances moved are influenced by temporal, geographical, demographic and mortality variables (see Table 1).

Methods

All puffin banding records for Newfoundland and Labrador between 1966 and 2001 (n = 8777) and all recoveries on the island between 1968 and 2001 (n=79) were furnished by the North American bird banding office. Newfoundland-banded birds that were recovered off the island were not available for analysis.[1] Thus this analysis has been performed on a subset of what is (hopefully) available. Of the 79 recovery records available, 10 were colony re-sightings of live birds, that were discarded from the analysis.

The following table shows the variables of interest in this study:

Table 1: Variables of interest

Banding date	Age at recovery
Banding location	Recovery season
Age at banding	Distance from colony when recovered
Recovery date	Cause of mortality
Recovery location	Area of Newfoundland where recovered – see Table 2 for list

Sample size varies between analyses due to the fact that not all pieces of information were available for all recovery records. This sometimes made analysis difficult particularly when interaction terms were of interest.

[1] I requested these but they did not arrive.

Age at banding and recovery is scored as either "adult" or "juvenile". A a bird banded in its first year of life is considered a juvenile. Young puffins begin to attend the breeding colony when they are three years old (Harris 1984) and thus a recovered bird is classified as juvenile (n=3), if it was known to be recovered in its first three years of life.

"Summer" is defined as 1 May – 15 Sept, the period when puffins attend the breeding colony. The rest of the year he is considered to be "winter".

Figure 1 shows banding and recovery locations for all birds in this data set. The banding locations are shown by "flags" and recovery locations by squares. The location of each recovery was used to assign it to one of four geographical recovery areas shown in Table 2.

Table 2: Recovery areas used in the analyses

Name of area	**Area Included**
Avalon Peninsula	Eastern half of Placentia and Trinity Bays, Conception Bay, St. Mary's Bay, eastern Avalon
Placentia Bay	Western half of Placentia Bay and Burin Peninsula
Northeast Coast	Western half of Trinity Bay to White Bay
Straits	Northern Peninsula and Strait of Belle Isle

Results

Most of the analyses that follow fall into two categories:

1. The straight line d*istance* between banding and recovery locations is examined using General Linear Models.

2. *Frequency counts* of recoveries are examined using Generalized Linear Models.

Residuals were analyzed graphically (where possible) for all analyses. No residual plots showed any arches or bowls, indicating that the structural models were appropriate. Many showed non-normality of errors (GLM) or heterogeneity of variance/deviance (GLM/GzLM), which is commented on individually in each analysis.

Figure 1: Banding and recovery locations of birds in this data set. Flags indicate banding locations; squares indicate recoveries.

Section 1

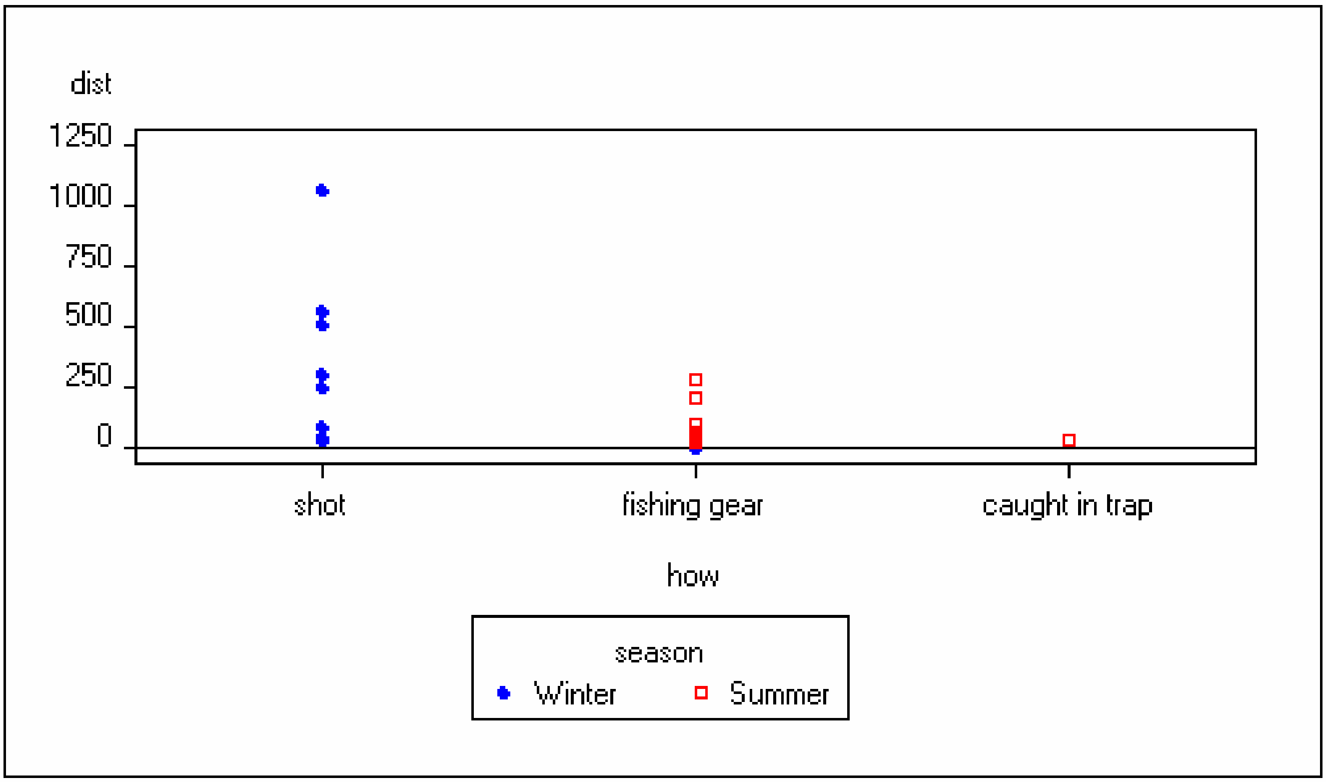
The first series of analyses looks at the response variable *distance*, the straight line distance between banding and recovery sites, in relation to a number of explanatory variables.

Distance From Colony When Recovered: Effect of Season and Cause of Death

I want to investigate the interactive effects of season and mortality type on recovery distance.

Verbal Model: does recovery distance due to a given cause of death depend on season

Graphical model:



Variables

Response: *dist* = distance from colony when recovered, on a ratio scale in kilometers.

Explanatory:

how = mortality type, on a nominal scale with three levels

season = recovery season, on a nominal scale with two levels

Formal model: $dist = \beta_o + \beta_{how} \cdot how + \beta_{season} \cdot season + \beta_{how \cdot season} \cdot how \cdot season + \varepsilon$

Hypothesis:

α = 5%

H_A: $\beta_{how \cdot season} \neq 0$ there is an interactive effect of mortality type and season on distance

H_o: $\beta_{how \cdot season} = 0$

Execution
Data in three columns, *dist, season* and *how*

```
MTB > GLM  'dist' = season how how * season;
SUBC>   Brief 3 .
```

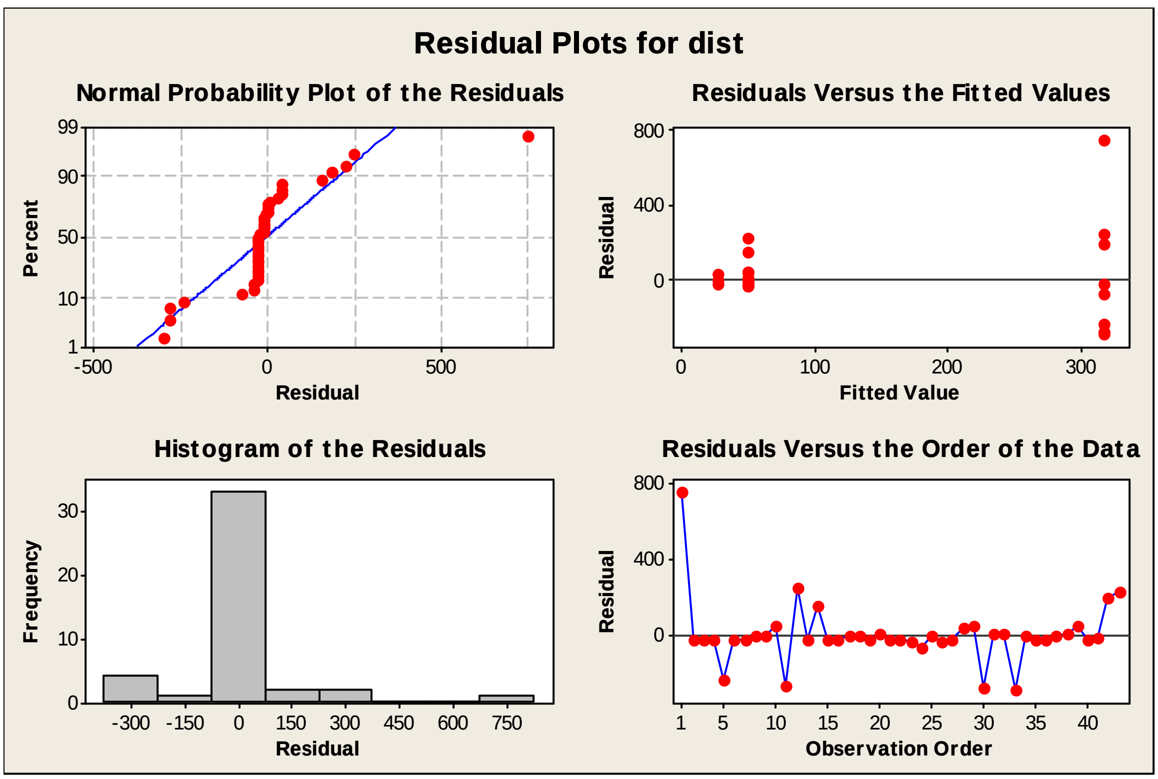
Since the data are unbalanced (eg. no records for shooting in summer) the interaction term causes the following error:

```
+ Rank deficiency due to empty cells, unbalanced nesting, collinearity, or an
     undeclared covariate. No storage of results or further analysis will be
     done.
```

So I removed the interaction term and ran the model again:

```
MTB > GLM  'dist' = season how;
SUBC>   Brief 3 .
```

Residual analysis



The residuals versus fits plot shows clear heterogeneity of variance. The normal probability plot and histogram of residuals show that the residuals are skewed and strongly strongly leptokurtic.

Results

```
Analysis of Variance for dist, using Adjusted SS for Tests

Source  DF   Seq SS   Adj SS  Adj MS     F      P
season   1   330557     1352    1352  0.05  0.823
how      1   189576   189576  189576  7.09  0.011
Error   40  1068930  1068930   26723
Total   42  1589063
```

Declare Decision

Interaction:
I cannot declare a decision about the test for the interaction term since the data were unbalanced and the term could not be included in the model.

Main effects:
The results of the analysis without the interaction term show that variation in recovery distance due to season is insignificant ($F_{1,40} = 0.05$, $p = 0.823$) when the significant variation due to cause of death ($F_{1,40} = 7.09$, $p = 0.011$) is taken into account.

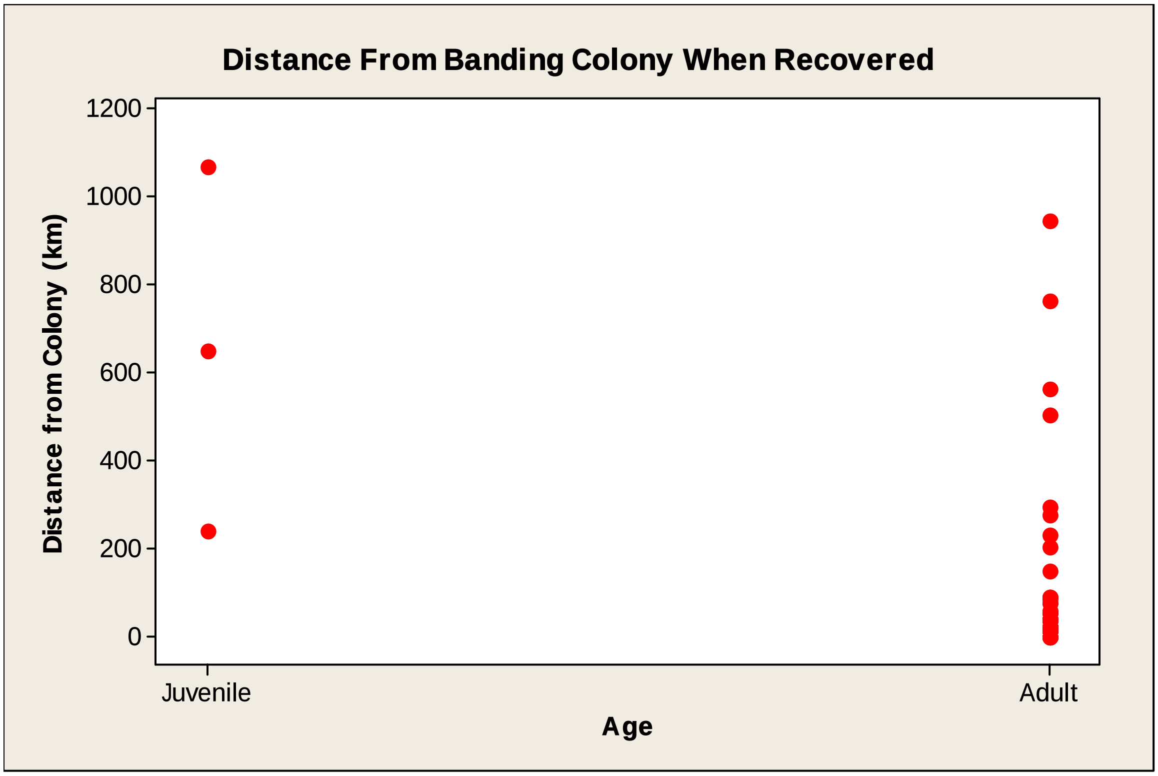
The p-value for mortality type is almost 5 times smaller than α and n is large, so re-computing the p-value by randomization is unlikely to change our decision.

Distance From Colony When Recovered: Effect of Age

I expect that juveniles moved further from the colony than adults

Verbal Model: is recovery distance greater for juveniles than for adults

Graphical model:



Variables

Response: *dist* = distance from colony when recovered, on a ratio scale in kilometers.

Explanatory: *age* = age when recovered, on a nominal scale with two levels (adult, juvenile)

Formal model: $\text{dist} = \beta_o + \beta_{age} \cdot \text{age} + \varepsilon$

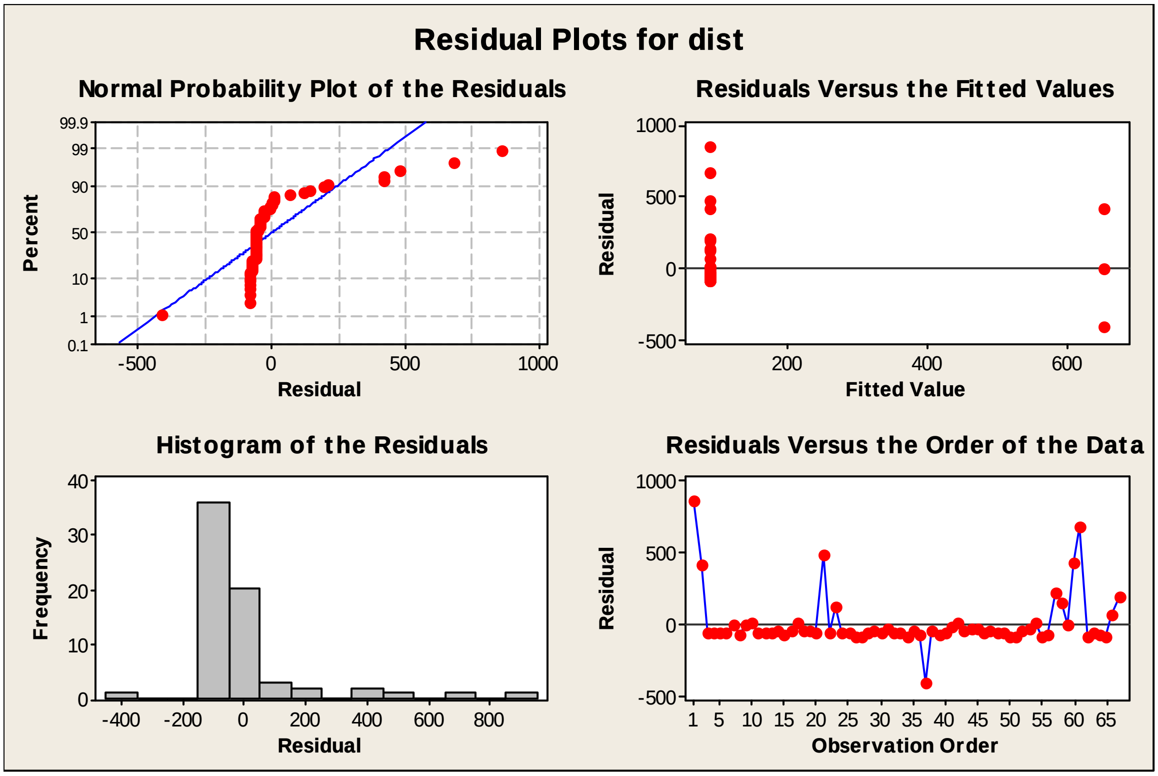
Hypothesis:
$\alpha = 5\%$

H_A: $\mu_{juvenile} > \mu_{adult}$ juveniles move further than adults
H_o: $\mu_{juvenile} = \mu_{adult}$

Execution
Data in two columns, *dist* and *age*

```
MTB > GLM  'dist' = age;
SUBC>   Brief 3 ;
SUBC>   GFourpack;
SUBC>   RType 1 .
```

Residual analysis



The residuals versus fits plot is difficult to interpret with only two columns of points, but the variance does appear relatively homogenous. The normal probability plot and histogram of residuals show that the residuals are strongly skewed and strongly leptokurtic.

Results

```
Analysis of Variance for dist, using Adjusted SS for Tests

Source  DF   Seq SS   Adj SS  Adj MS      F      P
age      1   917006   917006  917006  26.14  0.000
Error   65  2280311  2280311   35082
Total   66  3197317


S = 187.301   R-Sq = 28.68%   R-Sq(adj) = 27.58%


Term          Coef  SE Coef      T      P
Constant    371.24    55.32   6.71  0.000
age
adult      -282.84    55.32  -5.11  0.000
```

Declare Decision

Reject H_o, accept H_A that recovery distance depends on age ($F_{1,65} = 26.14$, $p < 0.0001$). Since this p-value is much smaller than α and n is large, the non-normality and heterogeneity of residuals will not affect our decision.

Analysis of Parameters
Means and standard errors of distance by age are given in the following table:

Descriptive Statistics: dist

Variable	age	Mean	SE Mean
dist	adult	88.4	21.9
	juvenile	654	239

The mean recovery distance for juveniles 654km (SE = 239) was significantly larger ($F_{1,65} = 26.14$, $p < 0.0001$) than that for adults 88.4 (SE = 21.9). This was expected since juveniles leave the their natal colony and are free to roam widely for three years, whereas adults are tied to the colony for part of each year

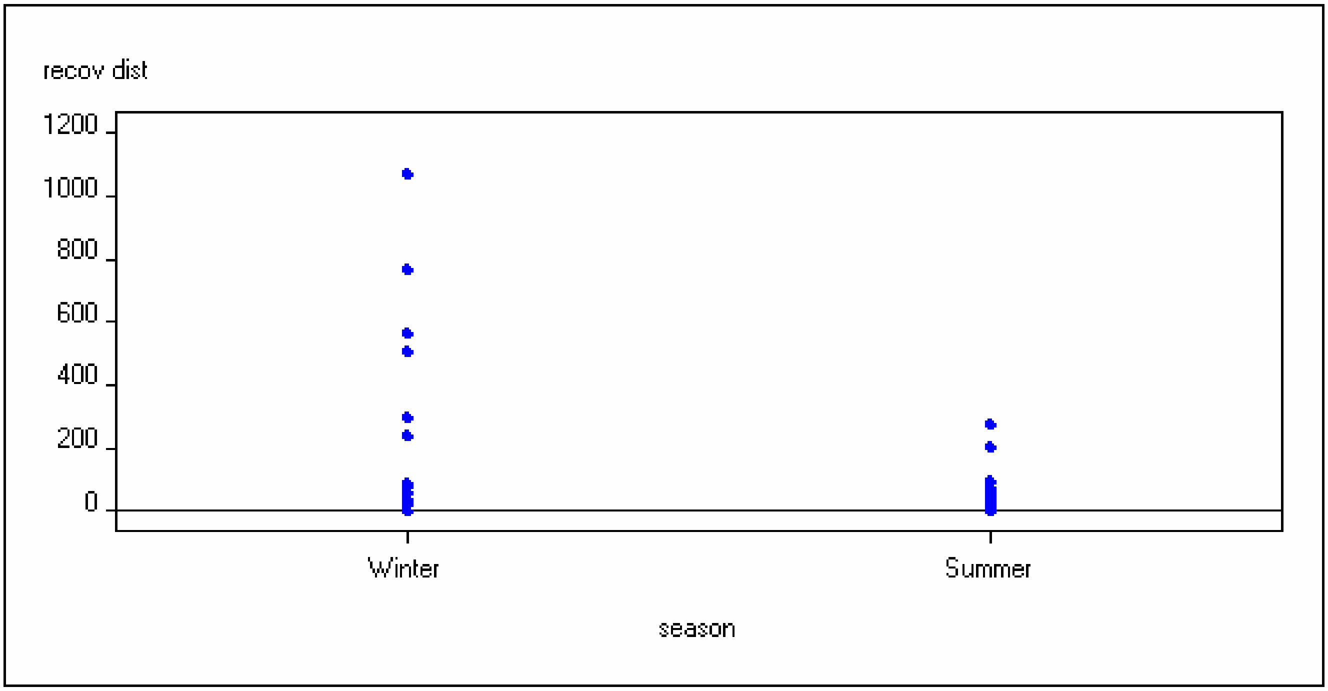
This analysis should be viewed with caution however since only three juveniles were included in the analysis versus 64 adults.

Distance from Colony When Recovered: Effect of Season

I want to test whether the mean recovery distance from the banding site depends on season of recapture i.e. a measure of movement by season. I expect that distances will be greater during winter when birds are not tied to the breeding colony.

Verbal Model: does recovery distance depend on season

Graphical model:



Variables

Response: *dist* = distance from colony when recovered, on a ratio scale in kilometers.

Explanatory: *season* = season when recovered, on a nominal scale with two levels (summer, winter)

Formal model: $dist = \beta_o + \beta_{season} \cdot season + \varepsilon$

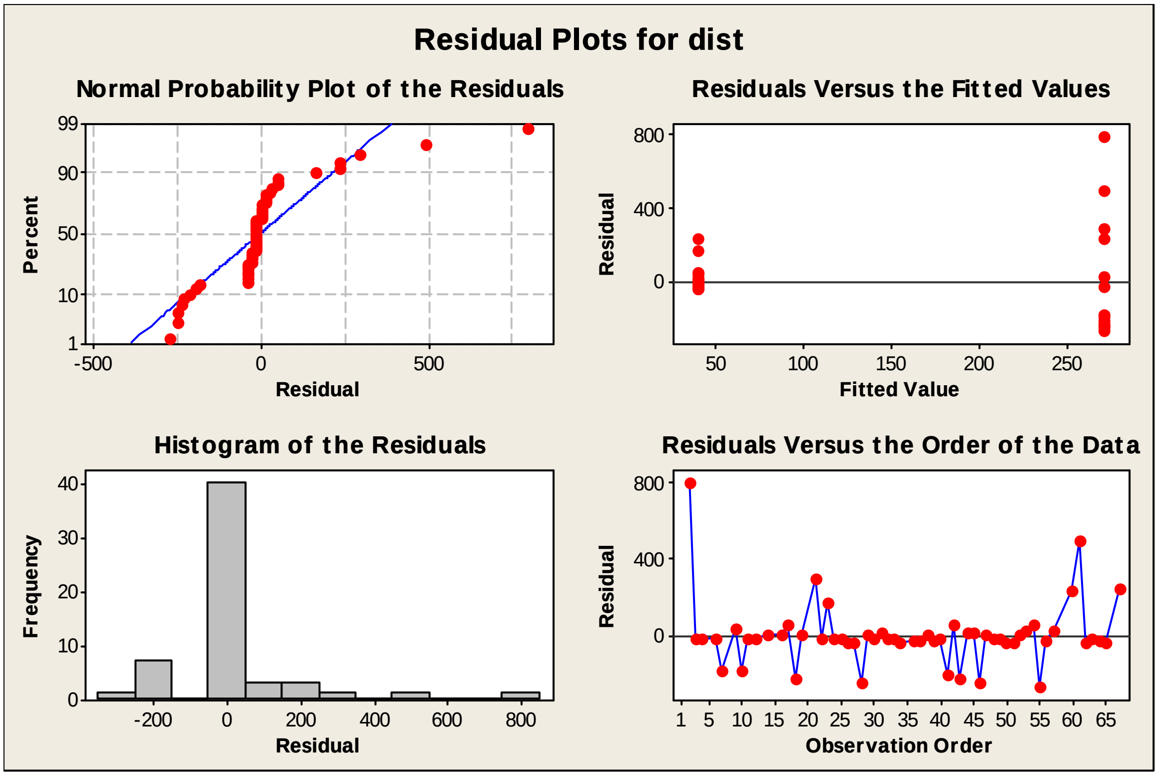
Hypothesis:
α = 5%

H_A: $\mu_{winter} > \mu_{summer}$ movement distance is greater in winter than summer
H_o: $\mu_{winter} = \mu_{summer}$

Execution
Data in two columns, *dist* and *season*

```
MTB > GLM  'dist' = season;
SUBC>   Brief 3 ;
SUBC>   GFourpack;
SUBC>   RType 1 .
```

Residual analysis



The residuals vs fits plot shows clear heterogeneity of variance. The normal probability plot and histogram of residuals show strongly leptokurtic and skewed residuals.

Results

```
Analysis of Variance for dist, using Adjusted SS for Tests

Source   DF    Seq SS    Adj SS  Adj MS      F      P
season    1    566694    566694  566694  19.94  0.000
Error    55   1563062   1563062   28419
Total    56   2129757

S = 168.580   R-Sq = 26.61%   R-Sq(adj) = 25.27%

Term         Coef  SE Coef      T      P
Constant   155.63    25.94   6.00  0.000
season
Summer    -115.82    25.94  -4.47  0.000
```

Declare Decision

Reject H_o, accept H_A that recovery distance depends on season ($F_{1,55} = 19.94$, $p < 0.0001$). Since this p-value is much smaller than α and n is large, the non-normality and heterogeneity of the residuals will not affect our decision.

Analysis of Parameters

Means and standard errors for distances by season are given in the following table:

Descriptive Statistics: dist

Variable	season	Mean	SE Mean
dist		211	103
	Summer	39.81	7.91
	Winter	271.4	89.3

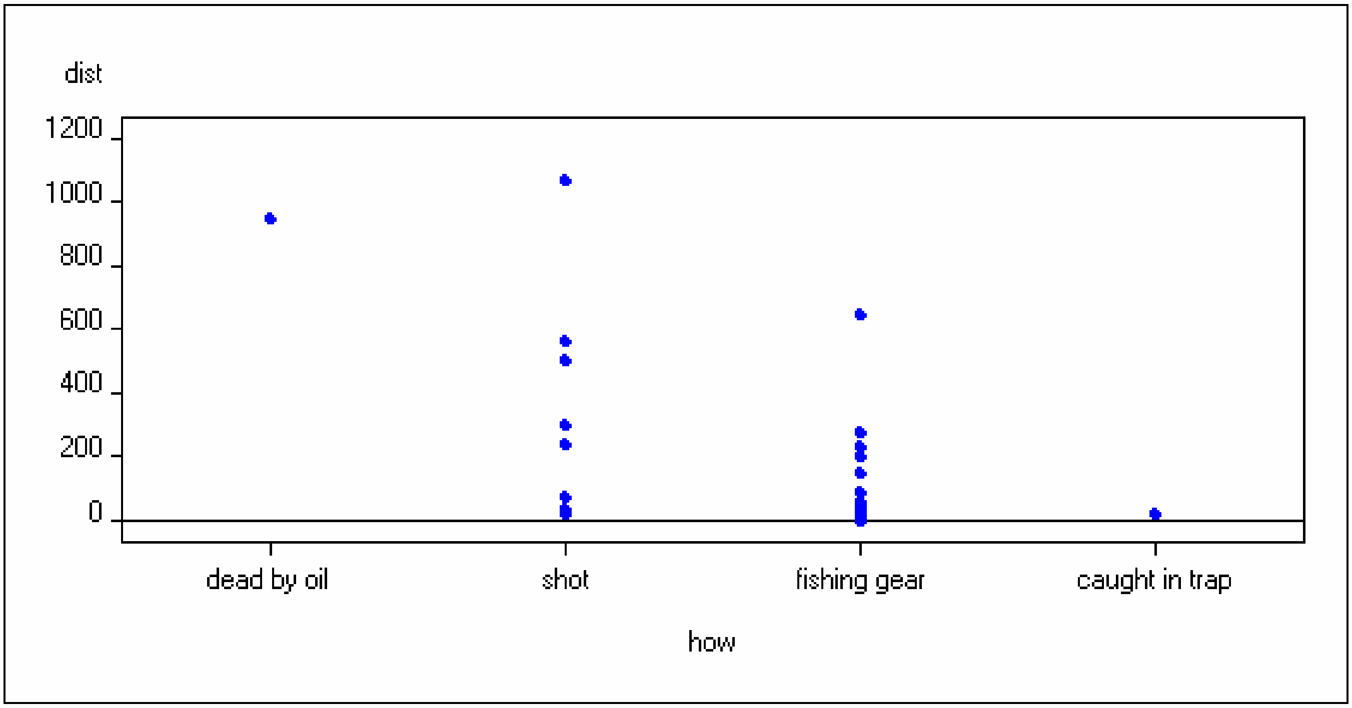
Birds are recovered within a mean distance of 39.81 (SE = 7.91) km of their banding site in summer compared to 271.4 (SE= 89.3) km in winter.

Distance from Colony When Recovered: Effect of Cause of Death

In this analysis I want to test whether different mortality risk factors operate at different distances from colony.

Verbal Model: does recovery distance depend on method of death

Graphical model:



Variables

Response: *dist* = distance from colony when recovered, on a ratio scale in kilometers.

Explanatory: *how* = how bird died, on a nominal scale with 4 levels (shot, fishing gear, caught in trap, dead by oil)

Formal model: $\text{dist} = \beta_o + \beta_{\text{how}} \cdot \text{how} + \varepsilon$

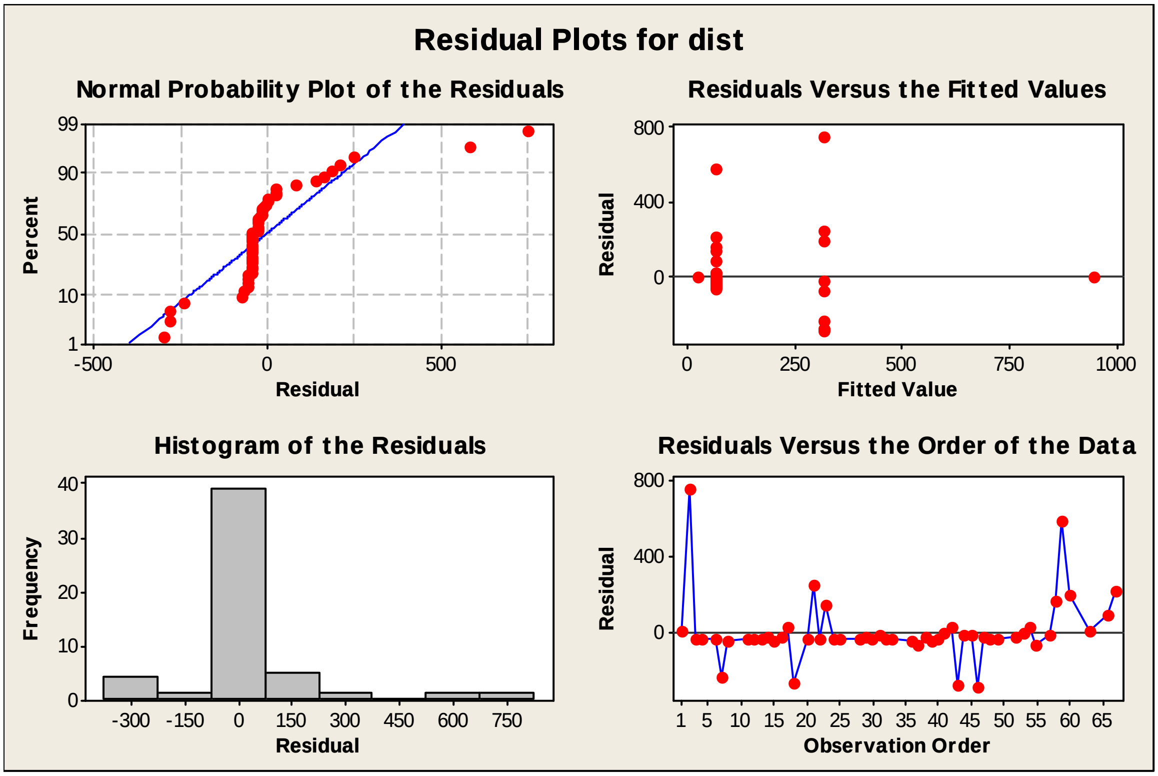
Hypothesis:
α = 5%

H_A: $\beta_{\text{how}} \neq 0$ there is variation in recovery distance due to method of death
H_o: $\beta_{\text{how}} = 0$

Execution
Data in two columns, *dist* and *how*

```
MTB > GLM  'dist' = how;
SUBC>   Brief 3 ;
SUBC>   GFourpack;
SUBC>   RType 1 .
```

Residual analysis



The residuals versus fits plot shows some indication of heterogeneity of variance. The normal probability plot and histogram of residuals show strongly leptokurtic and somewhat skewed residuals.

Results

```
Analysis of Variance for dist, using Adjusted SS for Tests

Source   DF   Seq SS    Adj SS   Adj MS      F      P
how       3  1163917   1163917   387972  12.71  0.000
Error    48  1465622   1465622    30534
Total    51  2629539


S = 174.739   R-Sq = 44.26%   R-Sq(adj) = 40.78%


Term                Coef  SE Coef      T      P
Constant          338.99    63.84   5.31  0.000
how
caught in trap    -316.6    139.1  -2.28  0.027
dead by oil        609.8    139.1   4.38  0.000
fishing gear     -272.30    66.69  -4.08  0.000
```

Declare Decision

Reject H_o, accept H_A that recovery distance depends on method of death ($F_{1,48} = 12.71$, $p < 0.0001$). Since this p-value is much smaller than α and n is large, the non-normality and heterogeneity of residuals will not affect our decision and randomization is not necessary.

Analysis of Parameters

Means and standard errors for distances by mortality type are given in the following table:

Descriptive Statistics: dist

Variable	how	Mean	SE Mean
dist		70.1	50.3
	caught in trap	22.437	*
	dead by oil	948.78	*
	fishing gear	66.7	17.4
	shot	318	116

Overall, birds are recaptured within 70 km (SE = 50.3) of their breeding colony. Birds tend to get caught by fishing gear close to their colony but get shot further away. The mean distance for entrapment in fishing gear is 66.7 (SE= 17.4) km whereas for shot it is considerably larger, 318 (SE=116) km.

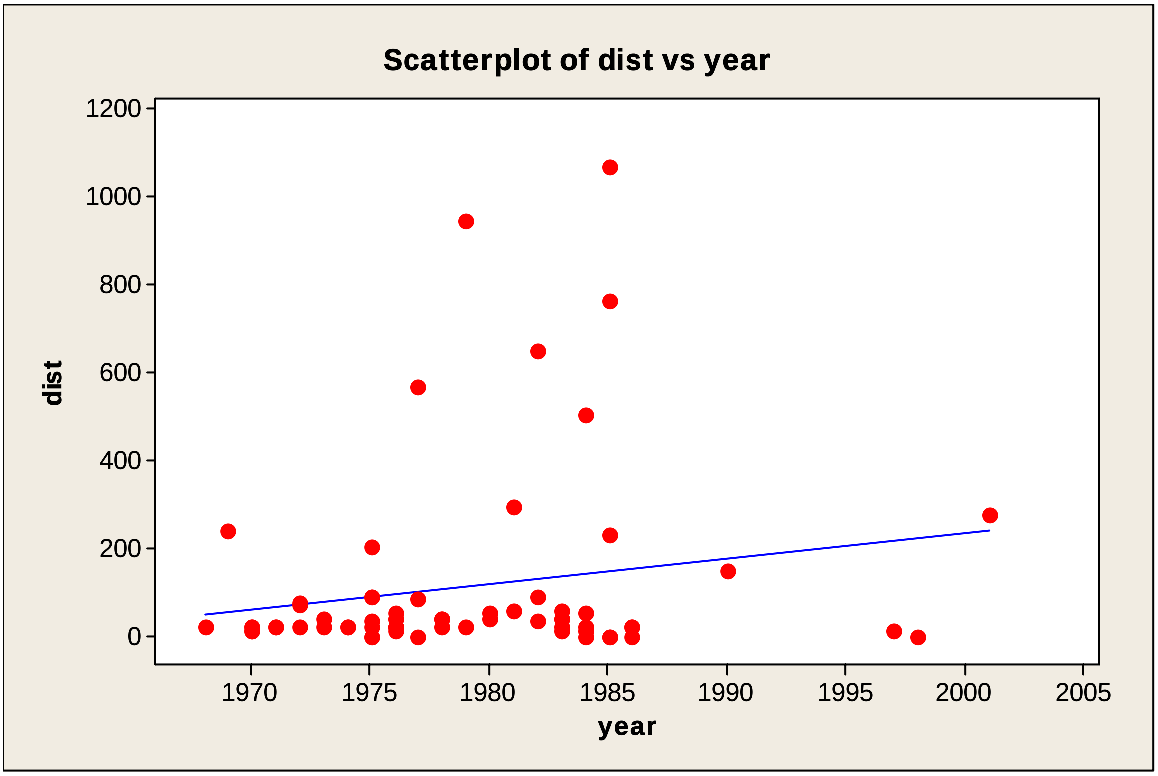
Not much can be said about "caught in trap" and "dead by oil" since there is only one recovery for each mortality type. Interestingly, the bird that died from oil was one of only two birds in the data set that were banded outside of Newfoundland – both from Maine; the other was shot.

Distance from Colony When Recovered: Effect of Year

I want to test whether there is the trend in recovery distance over time.

Verbal Model: does recovery distance depend on year

Graphical model:



Variables

Response: *dist* = distance from colony when recovered, on a ratio scale in kilometers.

Explanatory: *year* = year bird died, on an interval scale

Formal model: $dist = \beta_o + \beta_{year} \cdot year + \varepsilon$

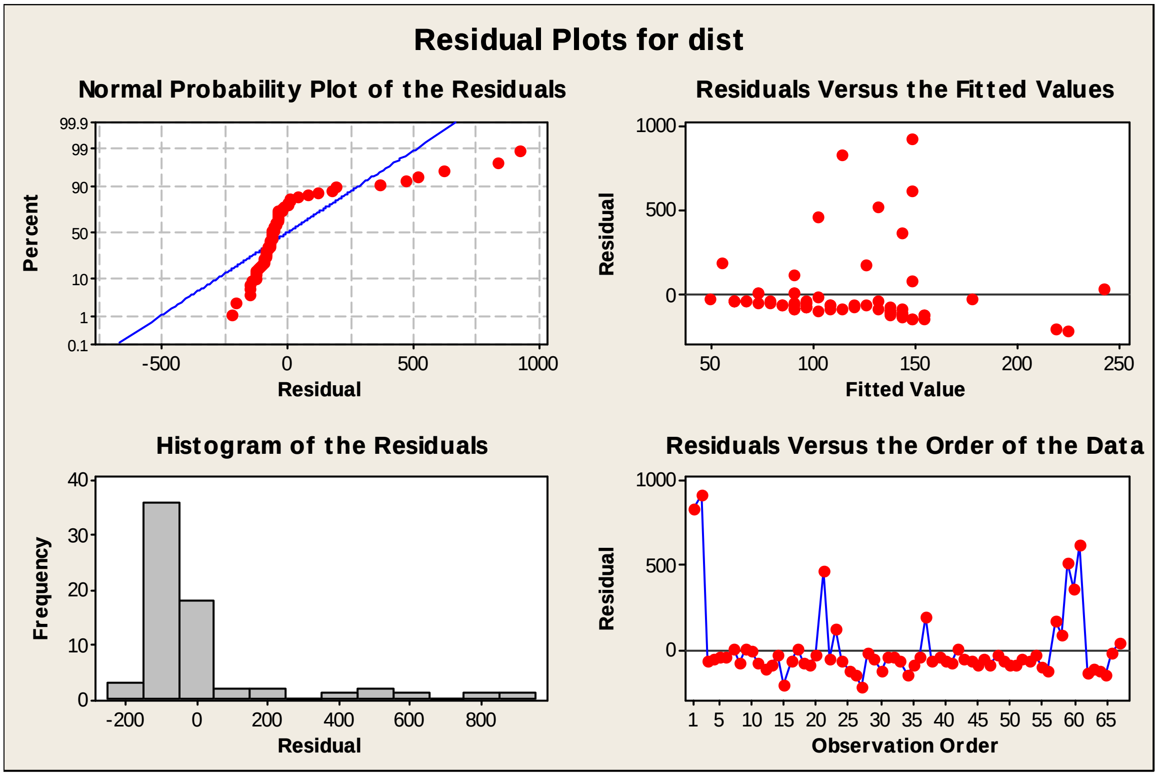
Hypothesis:
α = 5%

H_A: $\beta_{year} \neq 0$ recovery distance depends on year
H_o: $\beta_{year} = 0$

Execution
Data in two columns, *dist* and *year*

```
MTB > GLM  'dist' = year;
SUBC>   Covariates 'year';
SUBC>   Brief 3 ;
SUBC>   GFourpack;
SUBC>   RType 1 .
```

Residual analysis



The residuals versus fits plot shows heterogeneity of variance. The normal probability plot and histogram of residuals show that the residuals are strongly skewed and leptokurtic.

Results

```
Analysis of Variance for dist, using Adjusted SS for Tests

Source   DF    Seq SS    Adj SS   Adj MS     F      P
year      1    104220    104220   104220  2.19  0.144
Error    65   3093097   3093097    47586
Total    66   3197317
```

Declare Decision
Accept H_o that there is no relationship between recovery distance and year ($F_{1,66} = 2.19$, $p = 0.144$). The p-value is within a factor of three of α but n is large, so failure to meet the assumptions will not likely change our decision. If I were to publish this I would be paranoid and use randomization anyway.

Section 2

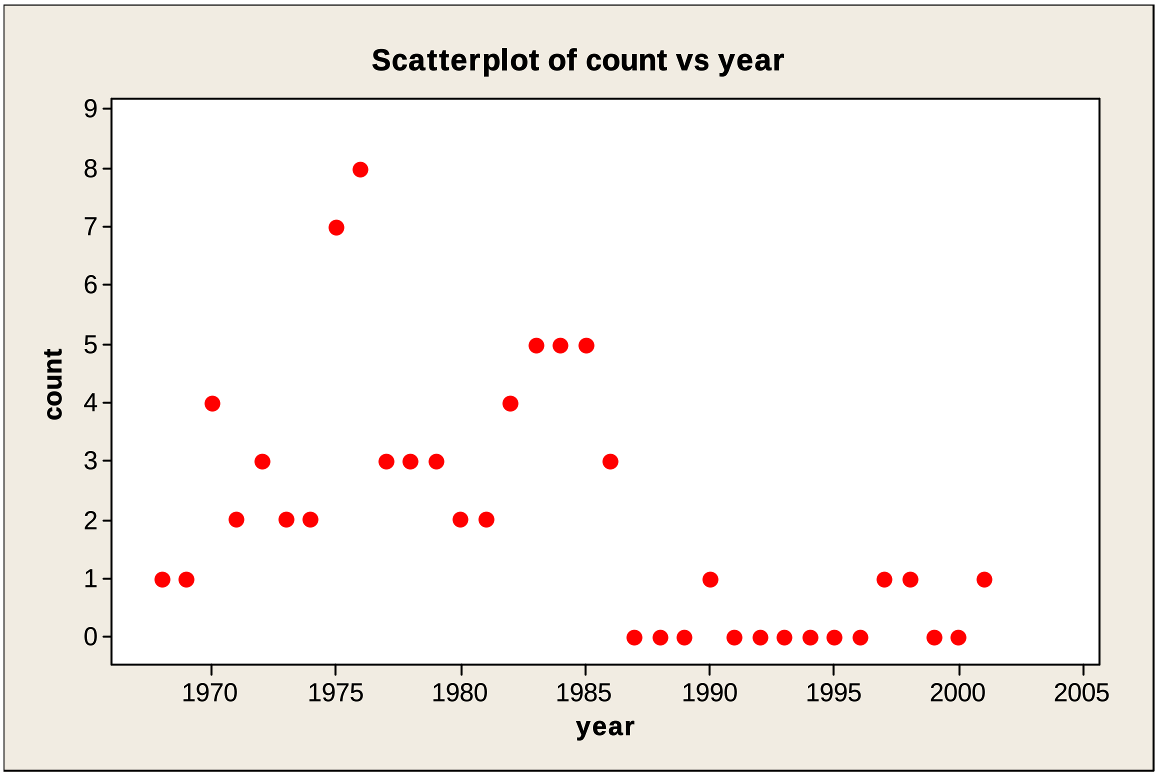
The next series of analyses looks at the frequency of recoveries, in relation to a number of explanatory variables using Generalized Linear Models

Yearly Trend in Band Returns

It appears that the number of band recoveries has decreased over time.
In this analysis I first try a General Linear Model and then a Generalized Linear Model.

Verbal Model: the number of band recoveries depends on year

Graphical model:



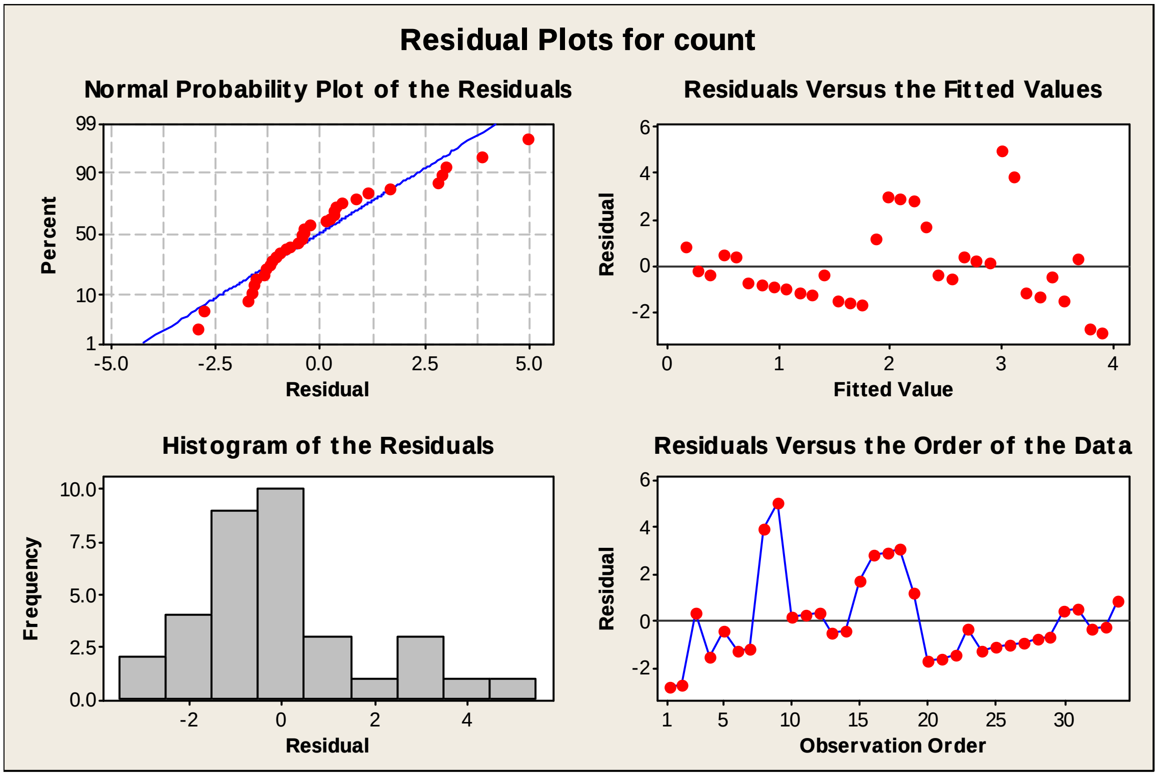
Formal model: $recoveries = \beta_o + \beta_{year} \cdot year + \varepsilon$

Execution

```
MTB > GLM 'recoveries' =  year;
SUBC>   Covariates 'year';
SUBC>   Brief 3 ;
SUBC>   GFourpack;
SUBC>   RType 1 .
```

Residual analysis

The following residual plots show heterogeneous variance and non-normal residuals thereby violating the assumptions for the General Linear Model. Therefore I'll start again with a Generalized Linear Model.



Generalized Linear Model

Treat data as counts with Poisson error structure.

Formal model:

$$\text{recoveries} = e^{u} + \varepsilon$$

$$u = \beta_o + \beta_{year} \cdot \text{year}$$

Hypothesis

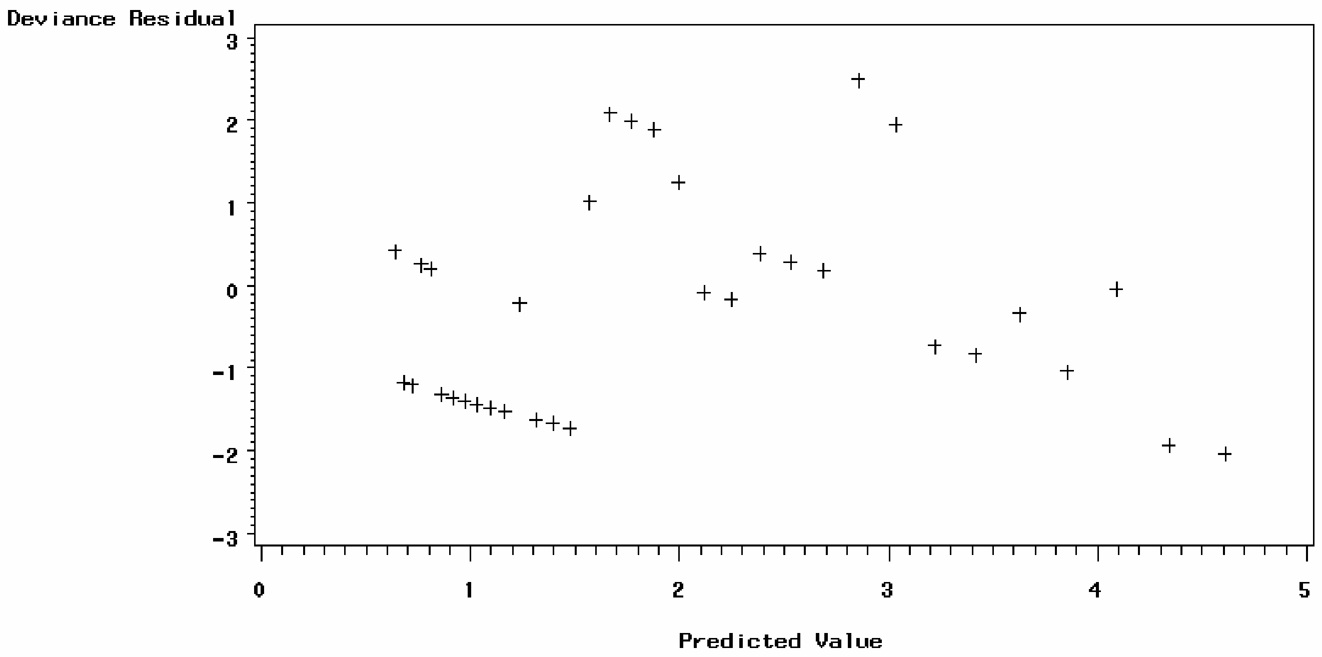
H$_A$: $\beta_{year} \neq 0$ there is a change in frequency of recovery by year.
H$_o$: $\beta_{year} = 0$

Execution

```
Proc genmod data=ATPU.count_by_year;
     model recoveries = year/ dist = poisson link = log type3;
   output out = genout pred=fits resdev=res;
run;
```

Residual analysis

The following residual vs. fits plot shows slightly better homogeneity than the one obtained using the General Linear Model indicating a better model fit.



Results

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	32	58.5594	1.8300
Scaled Deviance	32	58.5594	1.8300
Pearson Chi-Square	32	55.2704	1.7272
Scaled Pearson X2	32	55.2704	1.7272
Log Likelihood		-9.3802	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	119.3217	26.7787	66.8364	171.8070	19.85	<.0001
year	1	-0.0599	0.0135	-0.0864	-0.0333	19.57	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

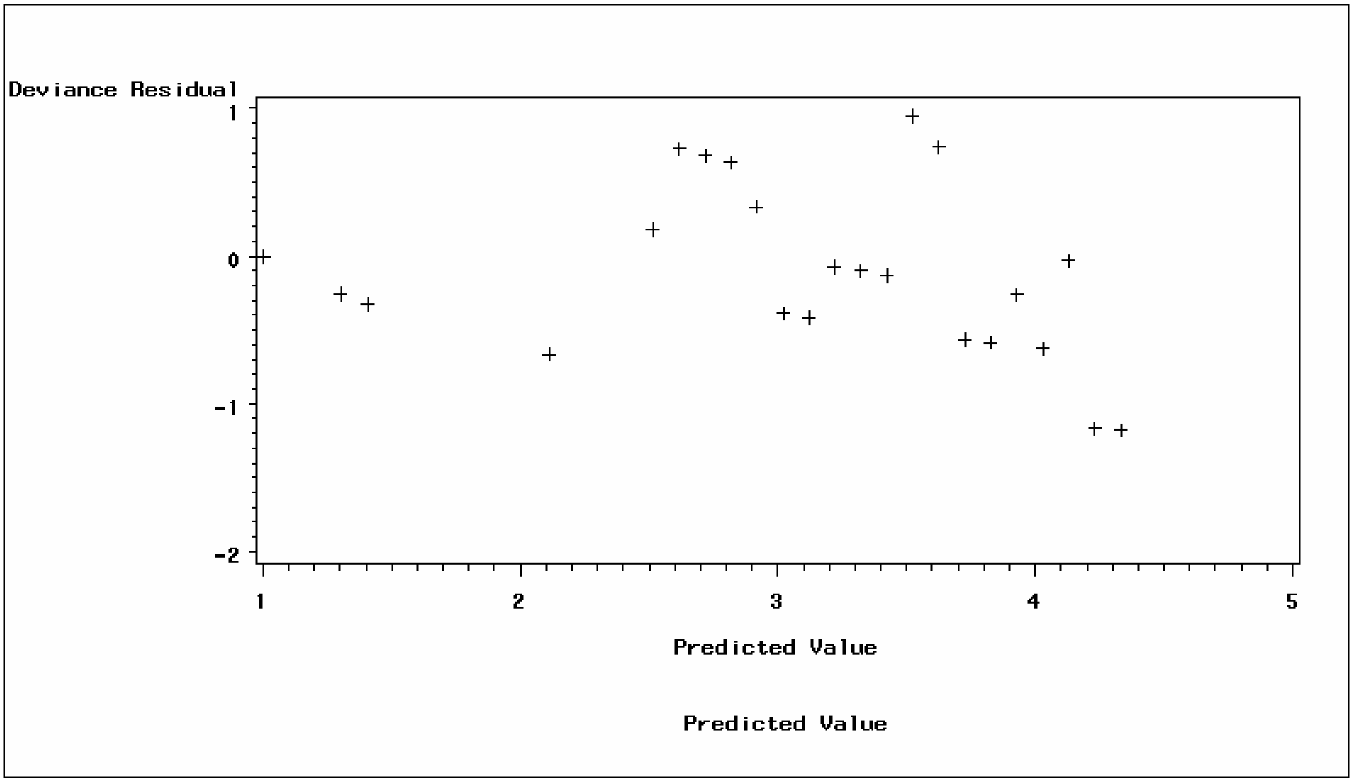
```
NOTE: The scale parameter was held fixed.


                 LR Statistics For Type 3 Analysis

                                             Chi-
                     Source          DF     Square    Pr > ChiSq
                     year             1      21.57        <.0001
```

Overdispersion

The "value/df" values highlighted in red in the preceding output indicate that the data are overdispersed. Two approaches are typically used to deal with this problem: 1) continue to use the Poisson distribution with the addition of a scaling factor (DSCALE option) or 2) use a different error distribution/link function (Littell et al. 2002; Stokes et al. 2000). I tried using the DSCALE option as well as the gamma and negative binomial distributions. The output and residual plots for each of these is shown below.

1) Reanalysis with dispersion correction using GENMOD DSCALE option:

```
                         Criteria For Assessing Goodness Of Fit

               Criterion                 DF           Value        Value/DF

               Deviance                  32         58.5594          1.8300
               Scaled Deviance           32         32.0000          1.0000
               Pearson Chi-Square        32         55.2704          1.7272
               Scaled Pearson X2         32         30.2027          0.9438
               Log Likelihood                       -5.1259

       Algorithm converged.


                            Analysis Of Parameter Estimates

                                        Standard     Wald 95% Confidence       Chi-
   Parameter     DF     Estimate         Error            Limits             Square     Pr > ChiSq

   Intercept      1     119.3217       36.2254      48.3212     190.3222      10.85         0.0010
   year           1      -0.0599        0.0183      -0.0957      -0.0240      10.69         0.0011
   Scale          0       1.3528        0.0000       1.3528       1.3528

NOTE: The scale parameter was estimated by the square root of DEVIANCE/DOF.


                           LR Statistics For Type 3 Analysis

                                                                Chi-
       Source         Num DF      Den DF     F Value     Pr > F      Square     Pr > ChiSq
       year                1          32       11.79     0.0017       11.79         0.0006
```



2) Re-analysis using gamma distribution with identity link function:

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	21	7.8148	0.3721
Scaled Deviance	21	24.2243	1.1535
Pearson Chi-Square	21	7.4002	0.3524
Scaled Pearson X2	21	22.9394	1.0924
Log Likelihood		-41.4386	

Algorithm converged.

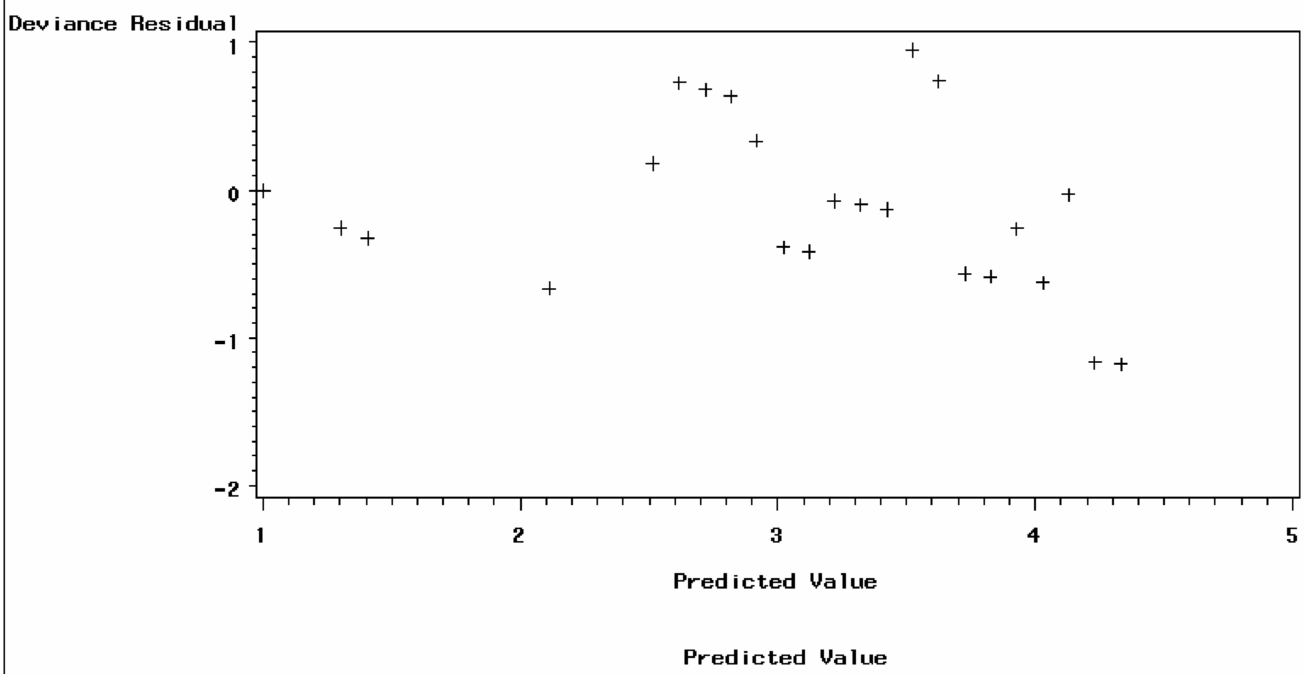
Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	202.8587	60.4960	84.2887	321.4286	11.24	0.0008
year	1	-0.1009	0.0304	-0.1605	-0.0413	11.00	0.0009
Scale	1	3.0998	0.8694	1.7890	5.3712		

NOTE: The scale parameter was estimated by maximum likelihood.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
year	1	4.41	0.0358



3) Re-analysis using negative binomial distribution with log link function:

Criteria For Assessing Goodness Of Fit

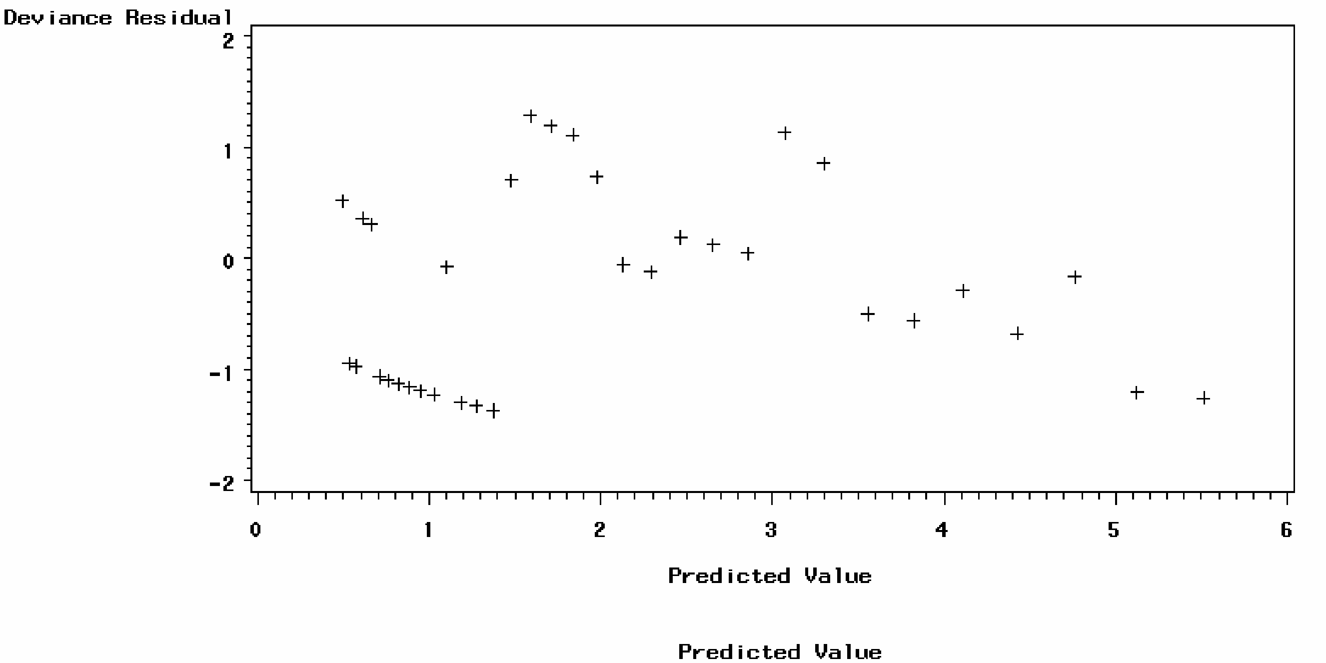
Criterion	DF	Value	Value/DF
Deviance	32	36.0929	1.1279
Scaled Deviance	32	36.0929	1.1279
Pearson Chi-Square	32	31.7851	0.9933
Scaled Pearson X2	32	31.7851	0.9933
Log Likelihood		-6.3422	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	145.6709	41.0775	65.1605	226.1814	12.58	0.0004
year	1	-0.0732	0.0207	-0.1138	-0.0325	12.45	0.0004

```
     Dispersion      1       0.3612       0.2269      -0.0835       0.8058

NOTE: The negative binomial dispersion parameter was estimated by maximum likelihood.


                         LR Statistics For Type 3 Analysis

                                              Chi-
                    Source           DF     Square     Pr > ChiSq

                    year              1      13.08         0.0003
```



Each of the three methods gives an improvement in the measure of dispersion (value/df in SAS output). To my eye, the residual plot for the negative binomial distribution looks best, but the differences between the three are subtle at best. The p-value differs between the three methods, ranging from 0.0003 (negative binomial) to 0.0358 (gamma). Other things being equal, an argument could be made for choosing the method with the most conservative p-value, i.e. 0.0358 using the gamma distribution in this case. The parameter estimates for the slope (ie. year term) generated by each method are quite similar; the 95% confidence interval for each method contains the slope parameter estimate generated by each of the other methods. I have chosen to report the p-value and parameter estimate obtained using the negative binomial distribution because the residual plot looks slightly better and the estimate of the slope parameter using the negative binomial lies between the estimates generated by the other two methods.

Declare Decision

Reject H_o, accept H_A that frequency of recovery depends on year ($G = 13.08$, $df = 1$, $p = 0.0003$).

Analysis of Parameters

Table 3 shows a comparison of the analyses for each method of over dispersion correction. Using the negative binomial method, for each unit increase in year, the number of band recoveries only changes by $e^{-0.0732}$ or 0.929. This means that for each unit increase in year, the number of bands recovered is only 92.9% (95% CI: 89.2% - 96.8%) what it was the previous year. This is equivalent to saying that there is a 7.1% (95% CI: 3.2% - 10.8%) annual decrease in band returns.

Table 3: Comparison of methods for correcting overdispersion. The method of choice was negative binomial.

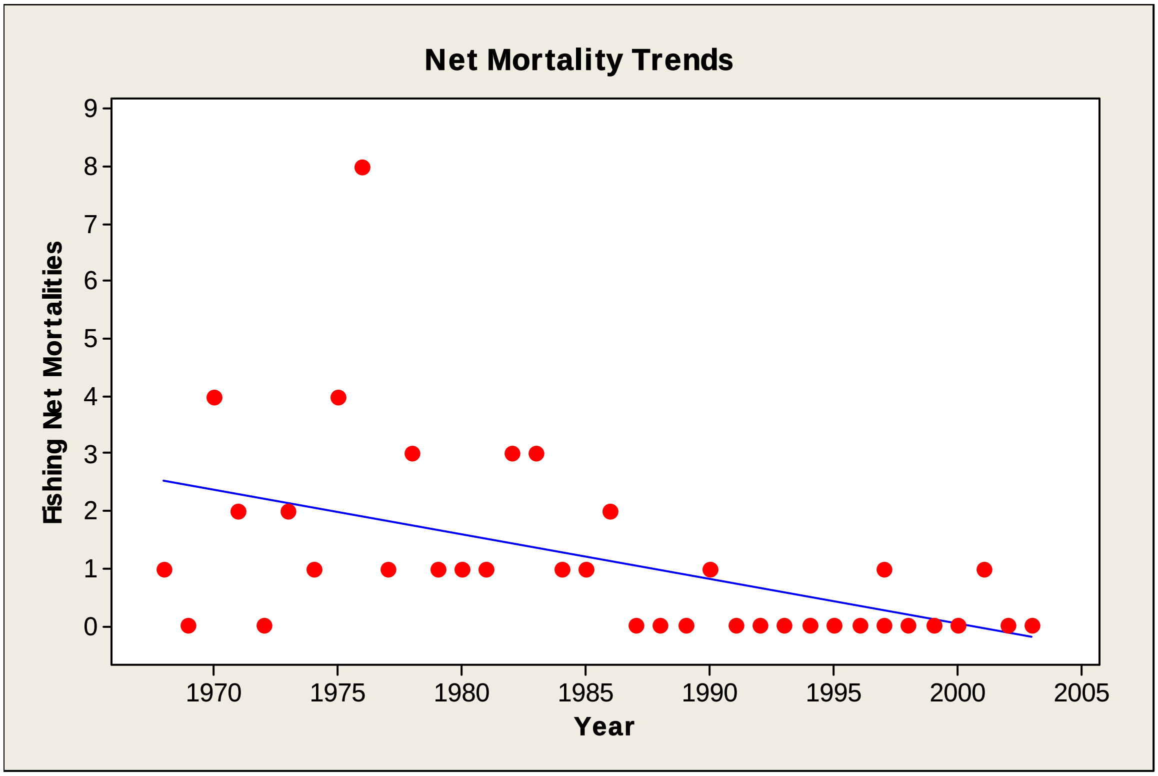
Distribution	Link	ΔDev	P-value	Parameter Estimate	Yearly % Decrease	95% CI (from Wald 95%)
Poisson	Log	21.57	<0.0001	$e^{-0.0599} = 0.942$	100 – 94.2 = 5.8%	3.3% - 8.3%
Poisson & DSCALE	Log	11.79	0.0017	$e^{-0.0599} = 0.942$	100 – 94.2 = 5.8%	2.4% - 9.1%
Gamma	Ident	4.41	0.0358	$e^{-0.1009} = 0.904$	100 – 90.4 = 9.6%	4.0% - 14.8%
Negative Binomial	**Log**	**13.08**	**0.0003**	$\mathbf{e^{-0.0732} = 0.929}$	**100 – 92.9 = 7.1%**	**3.2% - 10.8%**

Recovery Frequency: Netting Trends Across All Years

I suspect that the decreasing trend in band returns is driven by changes in fishing gear entanglement (netting) rates.

Verbal Model: netting frequency depends upon year

Graphical model:



Variables

Response: *count* = yearly number of recoveries due to net mortalities, on a ratio scale

Explanatory:

year = year, on an interval scale

Formal model:

$\text{count} = e^{u} + \varepsilon$

$u = \beta_o + \beta_{year} \cdot \text{year}$

Hypothesis

$\mathbf{H_A}$: $\beta_{year} \neq 0$ there is a trend in net mortalities over time – i.e. slope of regression line is not 0

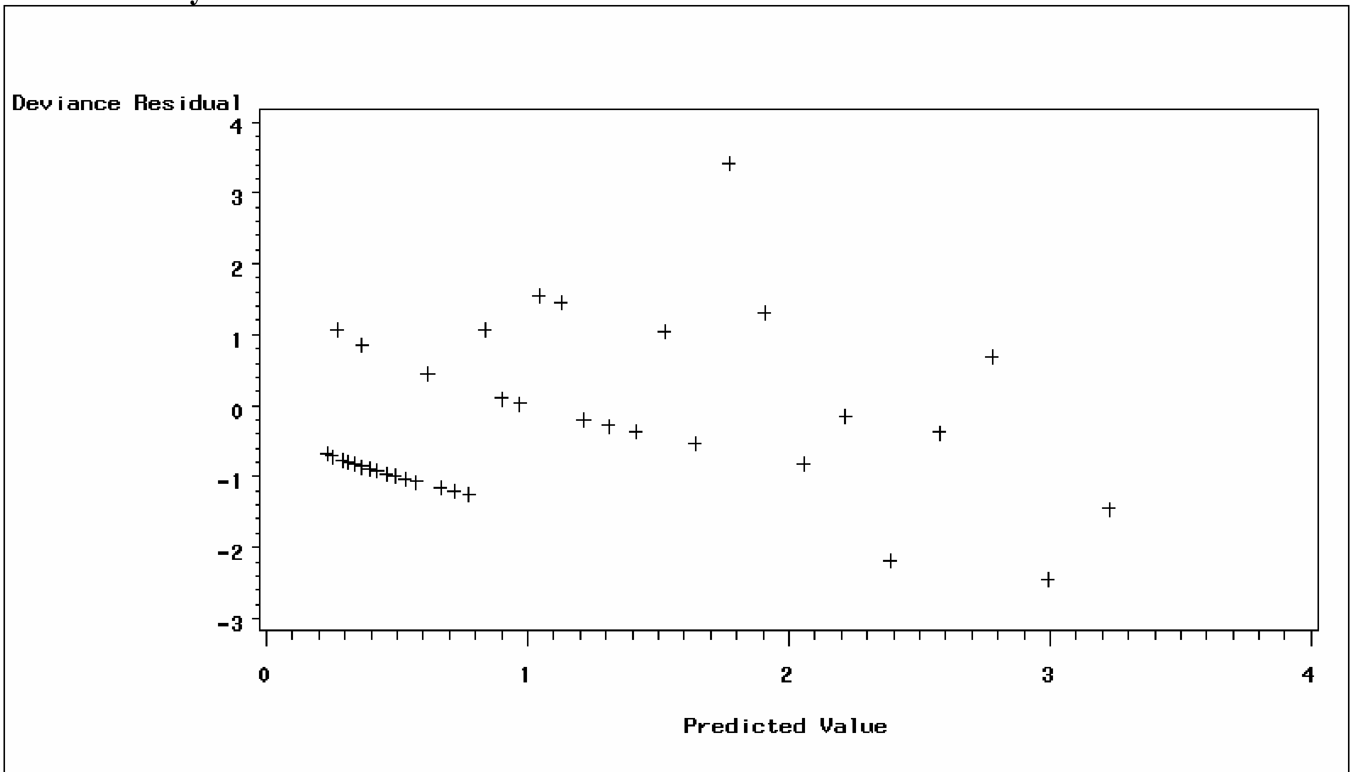
$\mathbf{H_o}$: $\beta_{year} = 0$

Execution

Data in two columns *count* and *year*

```
proc genmod data=ATPU.COUNT_fishing_year;
      model count = year / dist = poisson link = log type3;
      output out = genout pred=fits resdev=res;
run;
```

Residual analysis



The residuals versus fits plot shows reasonably homogenous residuals.

Results

These data display some overdispersion but since the "Value/DF" < 2, I judge that it is not severe enough to require reanalysis with an adjusted scale factor.

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	35	50.7351	1.4496
Scaled Deviance	35	50.7351	1.4496
Pearson Chi-Square	35	52.8004	1.5086
Scaled Pearson X2	35	52.8004	1.5086
Log Likelihood		-25.5953	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	148.8876	34.1479	81.9589	215.8164	19.01	<.0001
year	1	-0.0751	0.0173	-0.1089	-0.0412	18.91	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
year	1	22.16	<.0001

Declare decision
Reject $H_{o,}$ accept H_A that netting recovery frequency depends on year ($G = 22.16$, df=1, $p < 0.0001$).

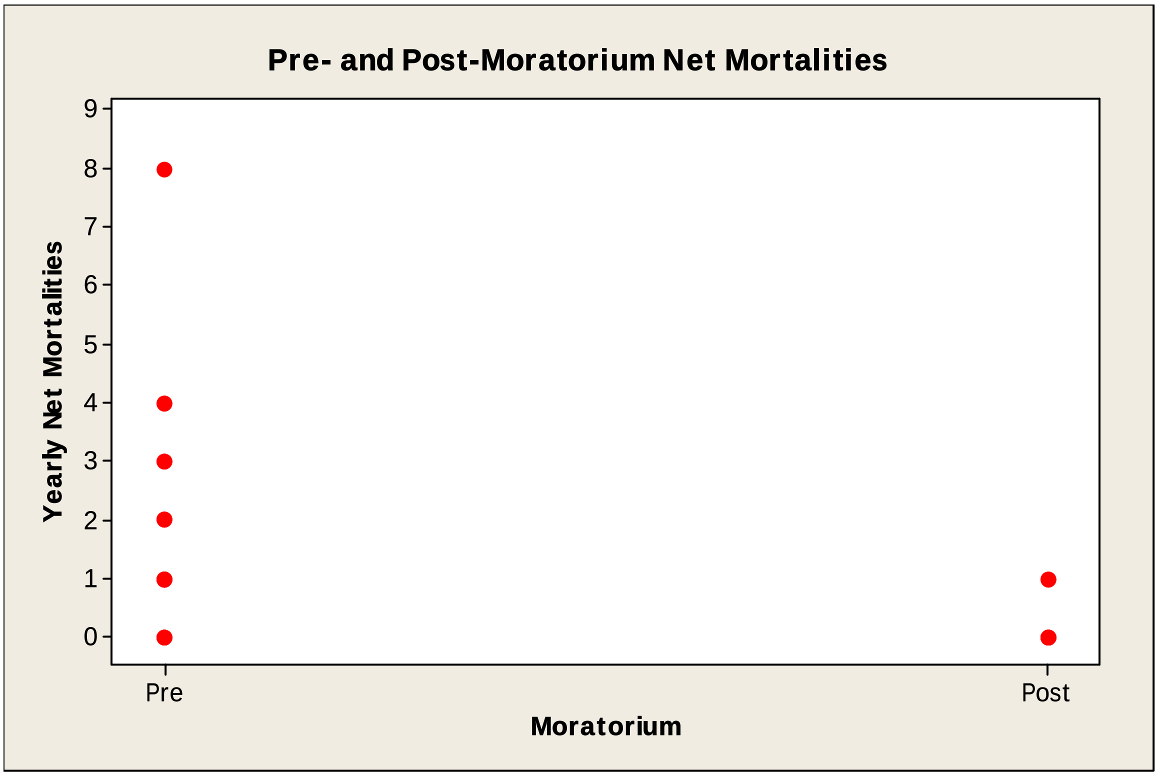
Analysis of parameters
The slope of the regression line is 0.9277 (i.e. $e^{-0.0751}$), (95% CI: 0.8968 - 0.9596). Thus the frequency of net mortality is decreasing by 7.23% (ie. 100% - 92.77%) yearly.

Recovery Frequency: Pre and Post Moratorium Netting Mortalities

I now want to investigate what is driving the long-term trend in netting mortalities. I suspect that the 1992 groundfish moratorium is responsible. Two questions arise: 1) is there a difference in mean nettings/year, pre-and post-moratorium (this analysis) and 2) when pre-and post-moratorium years are analyzed separately, are there any regression trends (next analysis) in each period.

Verbal Model: yearly net mortalities differ pre- and post-moratorium

Graphical model:



Variables

Response: *count* = yearly number of net mortalities

Explanatory:

mor = pre- or post-moratorium indicator, on a nominal scale with two values (pre, post)

Formal model:

count = $e^u + \varepsilon$

$u = \beta_o + \beta_{mor} \cdot \text{mor}$

Hypothesis

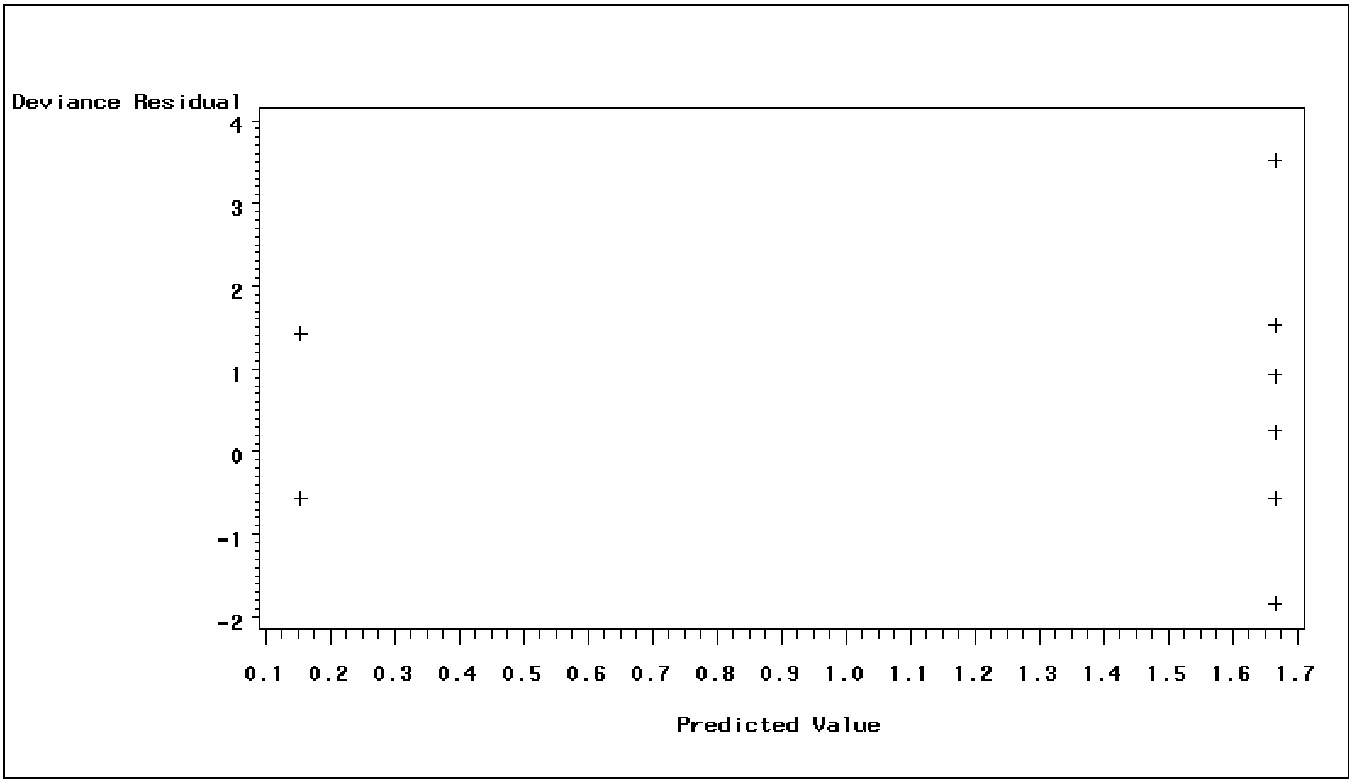
H_A: $\mu_{pre} > \mu_{post}$ Mean netting rates have decreased post-moratorium. (here μ is mean netting rate not the u in e^u in the model statement above).

H_o: $\mu_{pre} = \mu_{post}$

Execution
Data in two columns *count* and *mor*

```
proc genmod data=ATPU.COUNT_fishing_year;
      class mor;
      model count = mor / dist = poisson link = log type3;
      output out = genout pred=fits resdev=res;
run;
```

Residual analysis



The residuals versus fits plot shows heterogeneity of deviance.

Results
These data display some overdispersion but since the "Value/DF" < 2, I judge that it is not severe enough to require reanalysis with an adjusted scale factor.

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	35	50.1657	1.4333

```
               Scaled Deviance          35         50.1657          1.4333
               Pearson Chi-Square       35         57.4000          1.6400
               Scaled Pearson X2        35         57.4000          1.6400
               Log Likelihood                     -25.3106


       Algorithm converged.


                            Analysis Of Parameter Estimates

                                         Standard     Wald 95% Confidence       Chi-
Parameter              DF     Estimate       Error           Limits            Square    Pr > ChiSq

Intercept               1       0.5108      0.1581      0.2009      0.8207      10.44        0.0012
mor          post       1      -2.3826      0.7246     -3.8028     -0.9625      10.81        0.0010
mor          pre        0       0.0000      0.0000      0.0000      0.0000        .            .
Scale                   0       1.0000      0.0000      1.0000      1.0000


                           LR Statistics For Type 3 Analysis

                                                Chi-
                       Source           DF     Square    Pr > ChiSq
                       mor               1      22.73        <.0001
```

Declare decision

Reject $H_{o,}$ accept H_A that mean netting rates are significantly different pre- and post-moratorium ($G = 22.73$, df=1, $p < 0.0001$). Since n is large and the p-value is much smaller than α, the heterogeneity of residuals will not affect our decision and randomization is not necessary.

Analysis of parameters

The mean post-moratorium netting rate of $e^{(-2.3826+0.5108)}$ or 0.154 (SE=0.104) birds/year is 9.23% (95% CI: 2.22% - 38.19%) of the pre-moratorium rate of $e^{0.5108}$ or 1.667 (SE= 0.374) birds/year.

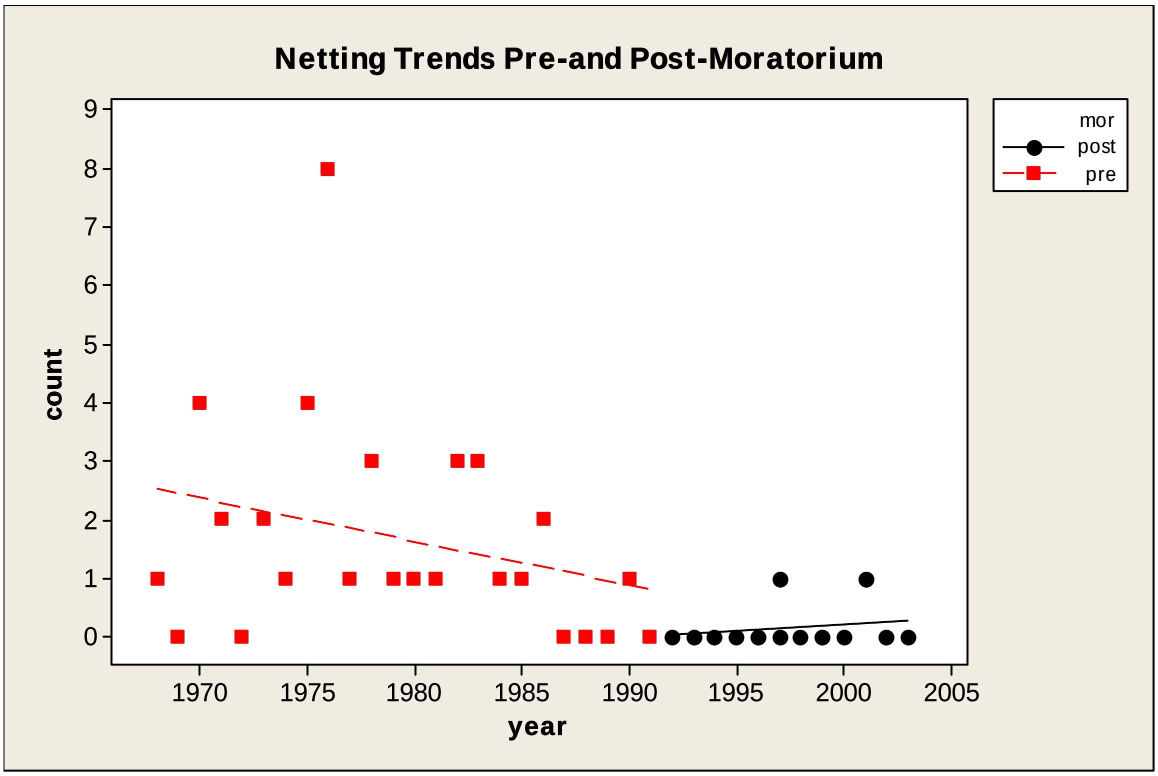
Recovery Frequency: Netting Trends Pre- and Post-moratorium

In the previous analysis I determined that the mean netting rates for pre- and post-moratorium years are significantly different. In this analysis I want to know whether there is a linear trend in netting rates when pre- and post-moratorium years are considered separately. This is accomplished with two separate Poisson regression analyses which are presented below.

Verbal Models:

netting frequency depends upon year, for pre-moratorium years
netting frequency depends upon year, for post-moratorium years

Graphical model:



Note: the regression lines drawn in this graph are not really representative of the analysis. They were drawn in Minitab using linear regression and do not take into account the log link function of the generalized linear model used.

Variables

Response: *count* = yearly number of net mortalities

Explanatory:

year = year, on an interval scale

Formal models:

count = $e^u + \varepsilon$

$u = \beta_o + \beta_{year} \cdot \text{year}$ for pre-moratorium years

$u = \beta_o + \beta_{year} \cdot \text{year}$ for post-moratorium years

Hypotheses

First analysis:
H_A: $\beta_{year} \neq 0$ there is a trend in net mortalities during pre-moratorium years
H_o: $\beta_{year} = 0$

Second analysis:
H_A: $\beta_{year} \neq 0$ there is a trend in net mortalities during post-moratorium years
H_o: $\beta_{year} = 0$

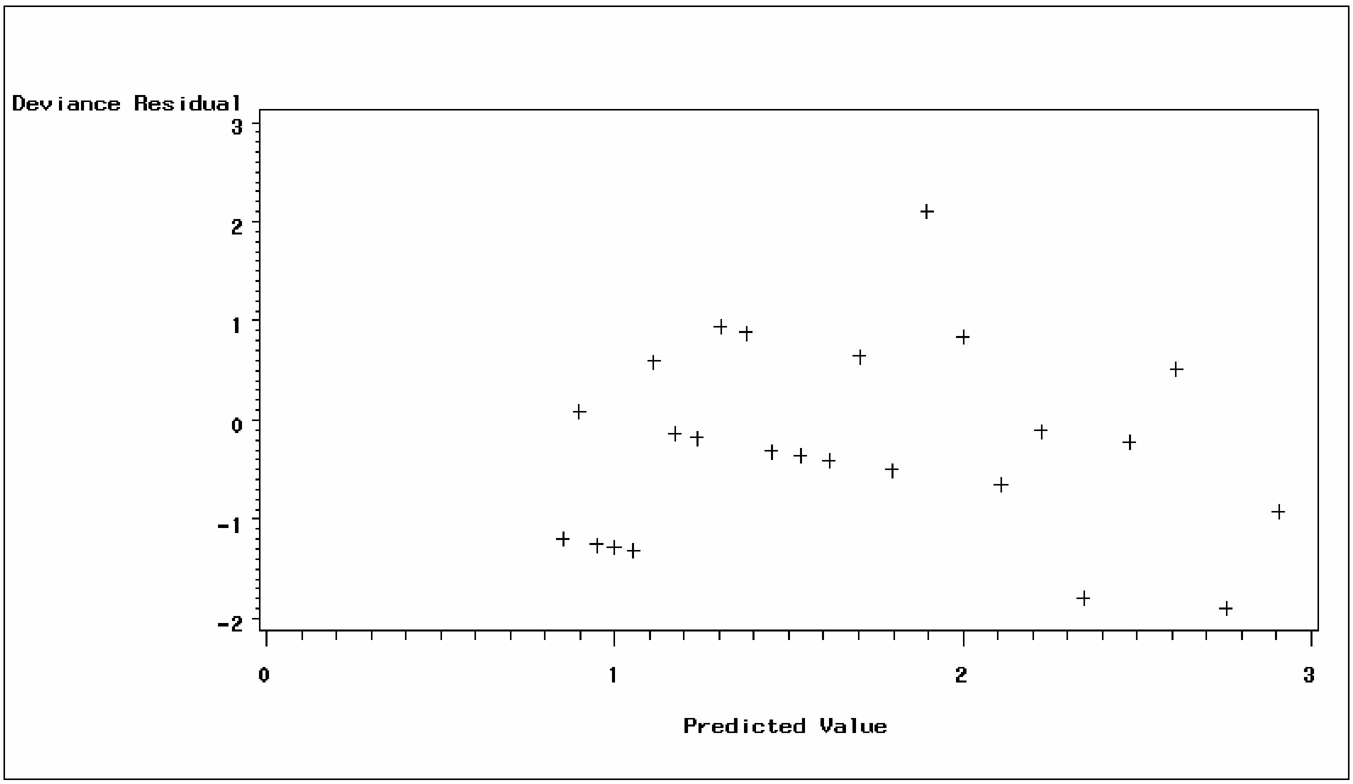
Execution:

First analysis: Pre-Moratorium Years
Data in two columns *count* and *year*

```
/*
 * Pre mor
 */
proc genmod data=pre;
      model count = year / dist = negbin link = log type3;
      output out = genout pred=fits resdev=res;
run;
```

Residual analysis

When I analyzed this dataset using the Poisson distribution, the data were overdispersed. I tried both the gamma and negative binomial distributions, and decided to use the latter as it gave the best residual plot.



The residuals versus fits plot shows minimal heterogeneity of variance.

Results

```
                         Criteria For Assessing Goodness Of Fit

              Criterion                 DF           Value        Value/DF

              Deviance                  22         24.6753          1.1216
              Scaled Deviance           22         24.6753          1.1216
              Pearson Chi-Square        22         23.9829          1.0901
              Scaled Pearson X2         22         23.9829          1.0901
              Log Likelihood                      -15.7546

                            Analysis Of Parameter Estimates

                                  Standard    Wald 95% Confidence       Chi-
     Parameter     DF    Estimate      Error          Limits           Square    Pr > ChiSq

     Intercept      1    106.1038    65.0937    -21.4775    233.6851     2.66        0.1031
     year           1     -0.0534     0.0329     -0.1179      0.0111     2.63        0.1048
     Dispersion     1      0.3704     0.2828     -0.1250      0.9248

NOTE: The negative binomial dispersion parameter was estimated by maximum likelihood.


                           LR Statistics For Type 3 Analysis

                                                   Chi-
                      Source            DF       Square    Pr > ChiSq
                      year               1         2.69        0.1010
```

Declare decision

Accept $H_{o,}$ reject H_A that recovery frequency depends on year (G = 2.69, df=1, p = 0.1010). I feel unsure whether the residual plot really shows any significant heterogeneity of the deviance residuals, and since n < 30 and the p-value is close to α , I would produce a p-value by randomization for publication.

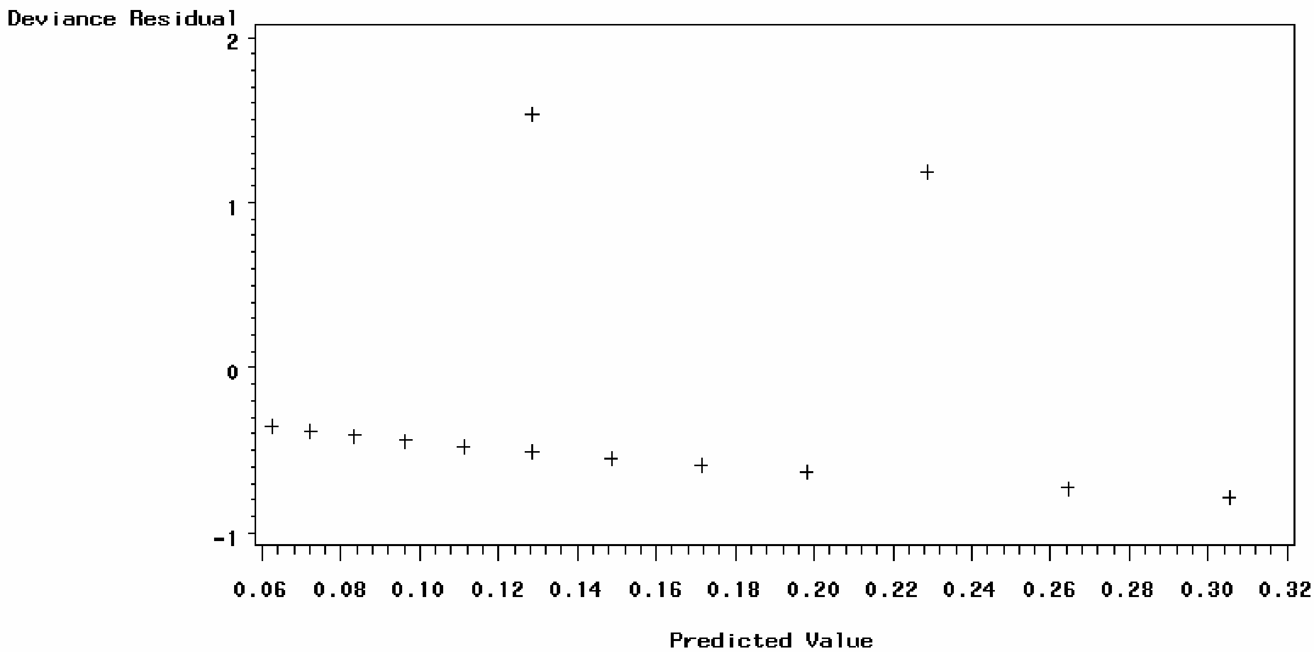
Execution:

Second analysis: Post-Moratorium Years

Data in two columns *count* and *year*

```
/*
 * post mor
 */
proc genmod data=post;
      model count = year / dist = poisson link = log type3;
      output out = genout pred=fits resdev=res;
run;
```

Residual analysis



The residuals versus fits plot shows homogenous residuals.

Results

There was no problem with overdispersion for this analysis so I used the Poisson distribution.

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	11	7.0513	0.6410
Scaled Deviance	11	7.0513	0.6410
Pearson Chi-Square	11	10.1458	0.9223
Scaled Pearson X2	11	10.1458	0.9223
Log Likelihood		-5.5257	

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-289.952	448.6613	-1169.31	589.4075	0.42	0.5181
year	1	0.1442	0.2244	-0.2957	0.5841	0.41	0.5207
Scale	0	1.0000	0.0000	1.0000	1.0000		

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
year	1	0.44	0.5091

Declare decision
Accept H_{o}, reject H_A that recovery frequency depends on year ($G = 0.44$, df=1, $p = 0.5091$).

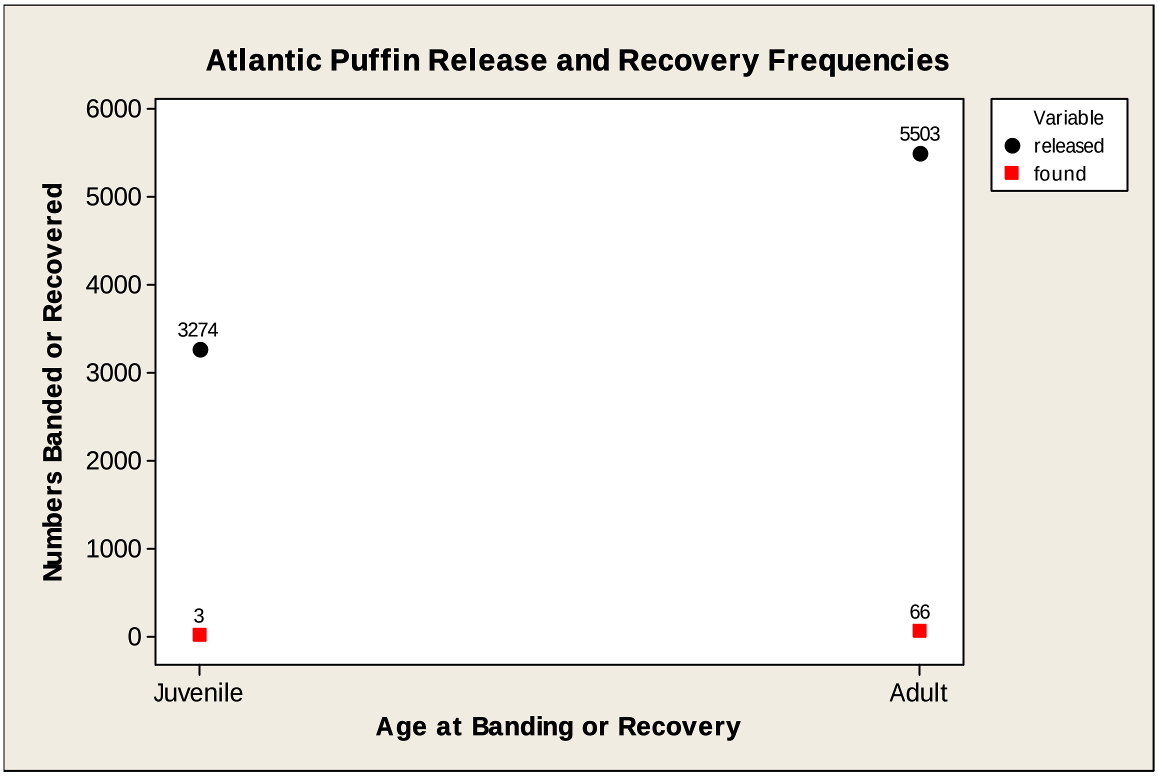
Analysis
When pre- and post-moratorium years are analyzed in isolation, no trend in netting rates is found for either period.

Age Structure of Recoveries and Bandings

For many organisms, the highest mortality occurs during the first year of life and this is true of puffins (Harris 1984). For this analysis, I compare the ratio of bandings and recoveries by age using logistic regression.

Verbal Model: the odds of recovery depend on age

Graphical model:



Variables

Response:

released = count of birds banded, on a ratio scale.
found = count of birds recovered, on a ratio scale.

Explanatory:

r_agecat = age of bird at recovery, on a nominal scale with two levels

Formal model:

$$\text{Odds} = e^{u} + \varepsilon$$

$$u = \beta_o + \beta_{r\_agecat} \cdot \text{r\_agecat}$$

Hypothesis:

H_A: $\beta_{r\_agecat} \neq 0$ i.e. odds is not a constant
H_o: $\beta_{r\_agecat} = 0$

Execution

Data in three columns, *found, released* and *r_agecat.*

```
MTB > BLogistic 'found' 'released' =  r_agecat;
SUBC>   ST;
SUBC>   Factors  'r_agecat';
SUBC>   Logit;
SUBC>   Reference  'r_agecat' 'juvenile';
SUBC>   Brief 3.
```

Residual analysis
Since this is a saturated model there are no residuals. Binomial error distribution is considered appropriate for binomial response variable.

Results

```
Logistic Regression Table

                                                Odds     95% CI
Predictor      Coef   SE Coef        Z      P  Ratio  Lower  Upper
Constant   -6.99424  0.577607   -12.11  0.000
r_agecat
 adult      2.58291  0.590733     4.37  0.000  13.24   4.16  42.13


Log-Likelihood = -381.531
Test that all slopes are zero: G = 43.113, DF = 1, P-Value = 0.000
```

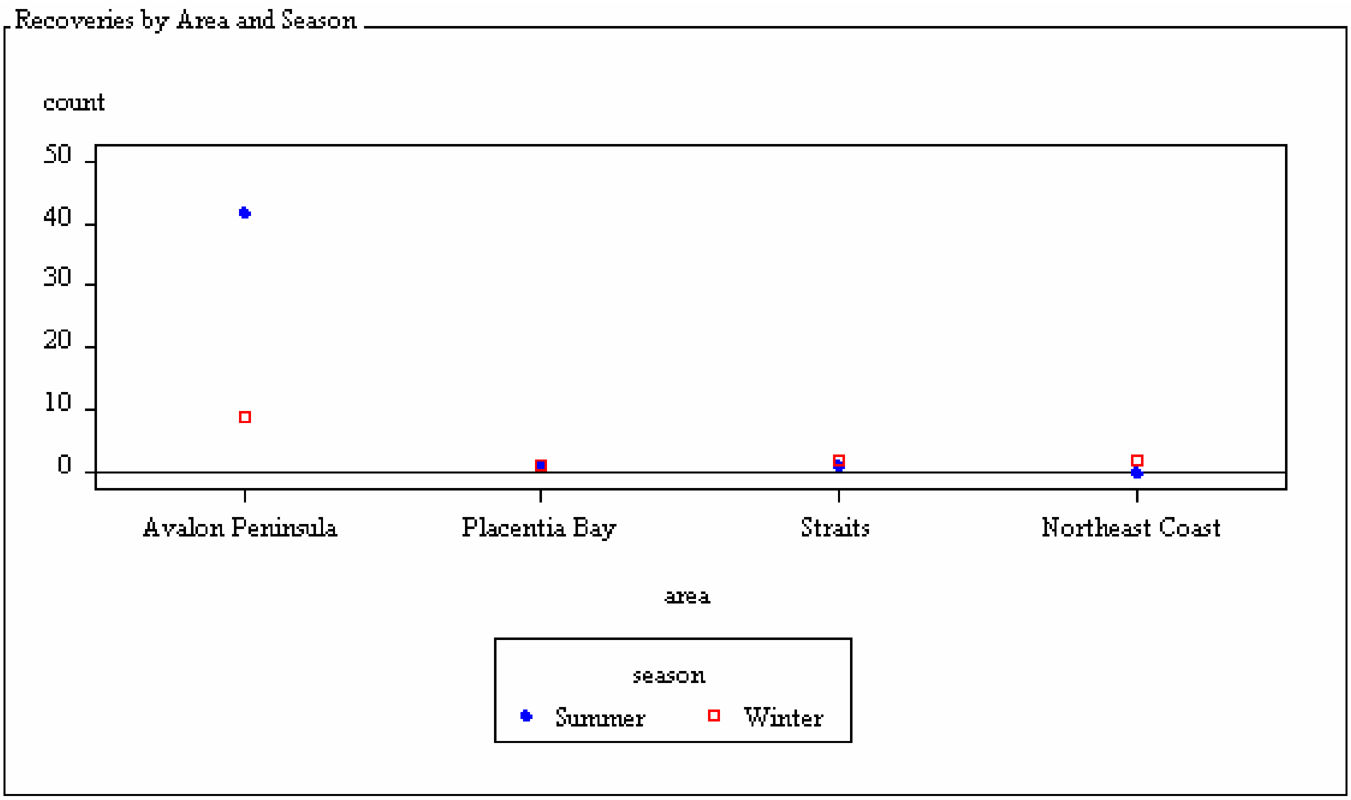
Declare Decision
The odds of recapture depend on age category ($G = 43.113$, $df = 1$, $p < 0.0001$), with adults being 13.24 (95% CI: 4.16 – 42.13) times more likely to be recaptured than juveniles. It is well-known from the literature that many more puffins die in the first few years of life than do adults (Harris 1984) and so we might naively expect the odds ratio to favor recovery of juveniles, but the opposite was found. Why? Juvenile puffins leave their breeding colonies before they are able to fly and swim out to sea. They do not return to land again until they are at least three years old, therefore, dead banded juveniles are likely to sink at sea rather than be found by humans.

Recovery Frequencies: Effects of Area and Season

I'm interested in knowing if the area where birds are recovered depends on season.

Verbal Model: area of recapture depends on season

Graphical model:



Contingency table:

area	Summer	Winter
Avalon Peninsula	42	9
Northeast Coast	0	2
Placentia Bay	1	1
Straits	1	2

Variables

Response: *count* = frequency of recovery by area and season

Explanatory:

area = area of recovery, on a nominal scale with four levels

season = time of year, on a nominal scale with two levels (summer, winter)

Formal model:

$$\text{count} = e^u + \varepsilon$$

$$u = \beta_o + \beta_{area} \cdot \text{area} + \beta_{season} \cdot \text{season} + \beta_{area \cdot season} \cdot \text{area} \cdot \text{season}$$

Hypothesis

Interaction:
$\mathbf{H_A}$: $\beta_{area \cdot season} \neq 0$ area of recovery depends on season
$\mathbf{H_o}$: $\beta_{area \cdot season} = 0$

Execution

Data in three columns, *count, area* and *season.*

```
proc genmod data=ATPU.count_season_area;
      class area season;
     model count = area|season/ dist = poisson link = log type3;
   output out = genout pred=fits resdev=res;
run;
```

Residual analysis
Since this is a saturated model, all residuals are zero.

Results

Analysis Of Parameter Estimates

Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept			1	0.6931	0.7071	-0.6928	2.0791	0.96
area	Avalon Peninsula		1	1.5041	0.7817	-0.0281	3.0363	3.70
area	Northeast Coast		1	0.0000	1.0000	-1.9600	1.9600	0.00
area	Placentia Bay		1	-0.6931	1.2247	-3.0936	1.7073	0.32
area	Straits		0	0.0000	0.0000	0.0000	0.0000	.
season	Summer		1	-0.6931	1.2247	-3.0936	1.7073	0.32
season	Winter		0	0.0000	0.0000	0.0000	0.0000	.
area*season	Avalon Peninsula	Summer	1	2.2336	1.2786	-0.2725	4.7397	3.05
area*season	Avalon Peninsula	Winter	0	0.0000	0.0000	0.0000	0.0000	.
area*season	Northeast Coast	Summer	1	-22.6931	84674.82	-165982	165936.9	0.00
area*season	Northeast Coast	Winter	0	0.0000	0.0000	0.0000	0.0000	.
area*season	Placentia Bay	Summer	1	0.6931	1.8708	-2.9736	4.3599	0.14
area*season	Placentia Bay	Winter	0	0.0000	0.0000	0.0000	0.0000	.
area*season	Straits	Summer	0	0.0000	0.0000	0.0000	0.0000	.
area*season	Straits	Winter	0	0.0000	0.0000	0.0000	0.0000	.
Scale			0	1.0000	0.0000	1.0000	1.0000	

Analysis Of Parameter Estimates

Parameter		Pr > ChiSq
Intercept		0.3270
area	Avalon Peninsula	0.0544
area	Northeast Coast	1.0000
area	Placentia Bay	0.5714

```
                    area          Straits                              .
                    season        Summer                               0.5714
                    season        Winter                               .
                    area*season   Avalon Peninsula   Summer            0.0807
                    area*season   Avalon Peninsula   Winter            .
                    area*season   Northeast Coast    Summer            0.9998
                    area*season   Northeast Coast    Winter            .
                    area*season   Placentia Bay      Summer            0.7110
                    area*season   Placentia Bay      Winter            .
                    area*season   Straits            Summer            .
                    area*season   Straits            Winter            .
                    Scale

                         LR Statistics For Type 3 Analysis
                                              Chi-
                      Source             DF     Square    Pr > ChiSq
                      area                3      57.31        <.0001
                      season              1       0.86        0.3529
                      area*season         3       9.99        0.0187
```

Declare decision

Reject $H_{o,}$ accept H_A that area of recovery depends on season (G = 9.99, df=3, p = 0.0187).

Analysis of parameters
Since the interaction term is significant I cannot analyze parameters for the main effects.

Reanalysis

I want to compare summer and winter recovery frequency in each of the four geographic regions. This focuses on *seasonal* differences within each *region*.

Formal Model:

For each of Avalon Peninsula, Northeast Coast, Placentia Bay, and Straits:

count = $e^u + \varepsilon$

$u = \beta_o + \beta_{season} \cdot$ season

Hypotheses

For Avalon Peninsula:
H_A: $\beta_{season} \neq 0$ frequency of recoveries on the Avalon Peninsula depends on season
H_o: $\beta_{season} = 0$

For the Northeast Coast:
H_A: $\beta_{season} \neq 0$ frequency of recoveries on the Northeast Coast depends on season
H_o: $\beta_{season} = 0$

For Placentia Bay:
H_A: $\beta_{season} \neq 0$ frequency of recoveries in Placentia Bay depends on season
H_o: $\beta_{season} = 0$

For the Straits:
H$_A$: $\beta_{season} \neq 0$ frequency of recoveries in the Straits depends on season
H$_o$: $\beta_{season} = 0$

Execution
Use the PROC GENMOD "by" option to perform separate analyses of each area:[2]

```
proc genmod data=ATPU.count_season_area;
      by area;
      class season;
    model count = season/ dist = poisson link = log type3;
run;
```

Residual analysis
Since these area saturated models, all residuals are zero.

Results

```
---------------------------------- area=Avalon Peninsula ----------------------------------
                             Analysis Of Parameter Estimates

                                     Standard   Wald 95% Confidence      Chi-
Parameter              DF   Estimate     Error         Limits          Square   Pr > ChiSq

Intercept               1     2.1972    0.3333     1.5439     2.8505    43.45       <.0001
season       Summer     1     1.5404    0.3673     0.8205     2.2604    17.59       <.0001
season       Winter     0     0.0000    0.0000     0.0000     0.0000      .           .
Scale                   0     1.0000    0.0000     1.0000     1.0000

                            LR Statistics For Type 3 Analysis

                                                Chi-
                        Source           DF     Square     Pr > ChiSq
                        season            1      23.17         <.0001

----------------------------------- area=Northeast Coast ----------------------------------

       WARNING: Negative of Hessian not positive definite.

The algorithm did not converge so no G statistic or p-value produced.

------------------------------------ area=Placentia Bay -----------------------------------

                             Analysis Of Parameter Estimates

                                     Standard   Wald 95% Confidence      Chi-
Parameter              DF   Estimate     Error         Limits          Square   Pr > ChiSq

Intercept               1     0.0000    1.0000    -1.9600     1.9600     0.00       1.0000
season       Summer     1     0.0000    1.4142    -2.7718     2.7718     0.00       1.0000
season       Winter     0     0.0000    0.0000     0.0000     0.0000      .           .
Scale                   0     1.0000    0.0000     1.0000     1.0000
```

[2]To convince myself that the "by" statement did what I thought, I created a data set containing cause of death for the Avalon timeslot only and analyzed it separately with PROC GENMOD. This produced identical results to those produced with the "by" statement.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
season	1	0.00	1.0000

------------------------------------ area=Straits ------------------------------------

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	0.6931	0.7071	-0.6928	2.0791	0.96	0.3270
season	Summer	1	-0.6931	1.2247	-3.0936	1.7073	0.32	0.5714
season	Winter	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
season	1	0.34	0.5599

Avalon Peninsula

When the Avalon Peninsula is analyzed separately, frequency of recovery depends on season (G = 23.17, df=1, $p < 0.0001$). Recoveries during summer are 4.7 (or $e^{1.5404}$) times (95% CI: 2.27 - 9.59) more frequent than recoveries during winter.

The Northeast Coast

On the Northeast Coast, there were no recoveries during summer so the algorithm failed to converge.

Placentia Bay

When Placentia Bay is analyzed separately, frequency of recovery does not depend on season (G = 0.00, df=1, p = 1.00) since there were the same number of recoveries (1) in each season.

Straits

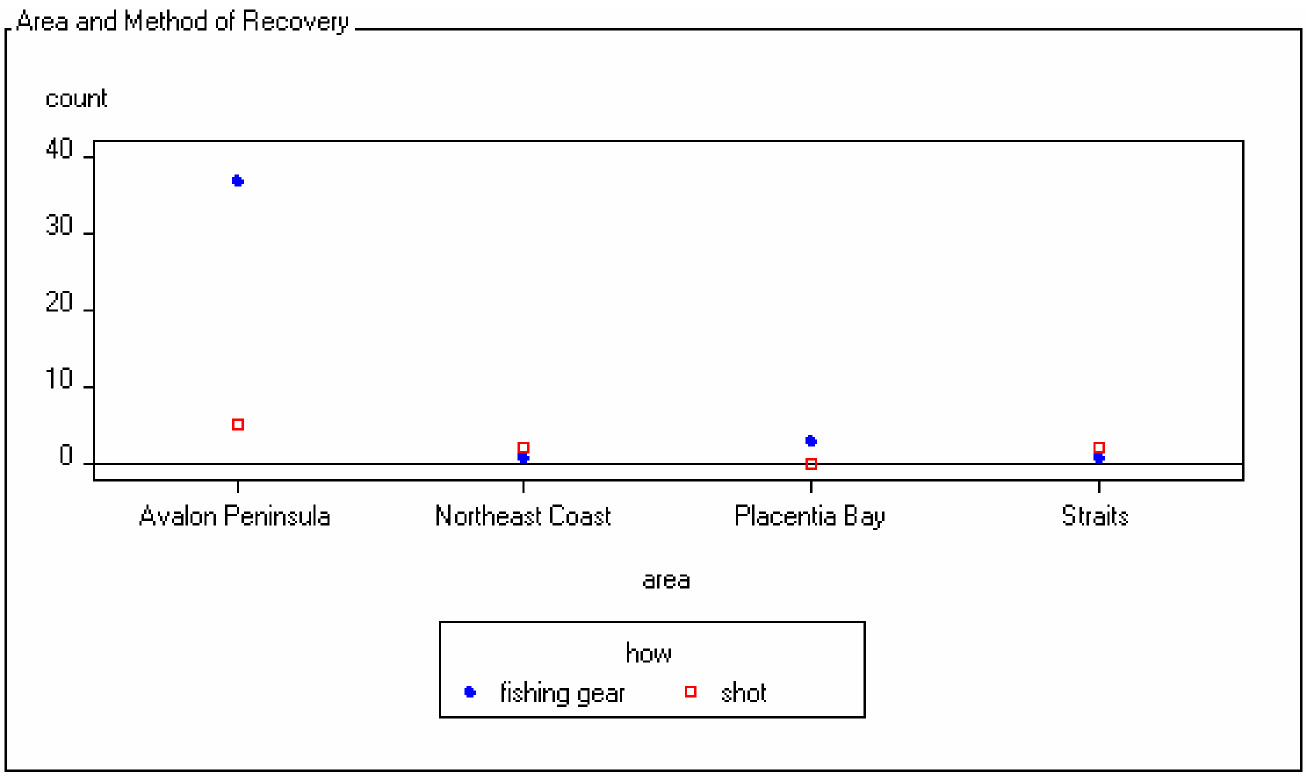
When the Straits is analyzed separately, frequency of recovery does not depend on season (G = 0.34, df=1, p = 0.5599) since there were the same number of recoveries (1) in each season.

Recovery Frequencies: Effects Of Cause of Death and Location

I want to investigate whether cause of death (restricted to fishing gear vs. shot) depends on area found.

Verbal Model: method of death depends on area found

Graphical model:



Contingency Table

Area	fishing gear	shot
Avalon Peninsula	37	5
Northeast Coast	1	2
Placentia Bay	3	0
Straits	1	2

In my first attempt at this analysis I had three levels for *how*: fishing gear, shot and other. This produced a contingency table with too many zeros and the maximum likelihood algorithm failed to converge. Removing the "other" category allowed the algorithm to converge. This makes more biological sense since the "other" category didn't really tell me anything biologically interesting.

Variables

Response: *count* = frequency of recovery by area and method of death

Explanatory:

area = area of recovery, on a nominal scale with four levels
how = cause of death, on a nominal scale with two levels (fishing gear, shot)

Formal model:

count = $e^u + \varepsilon$

$u = \beta_o + \beta_{area} \cdot \text{area} + \beta_{how} \cdot \text{how} + \beta_{how \cdot area} \cdot \text{how} \cdot \text{area}$

Hypothesis:

Interaction:
H_A: $\beta_{how \cdot area} \neq 0$ type of death depends on area found
H_o: $\beta_{how \cdot area} = 0$

Main effect:
H_A: $\beta_{how} \neq 0$ recovery frequency depends on type of death
H_o: $\beta_{how} = 0$

Execution

Data in three columns, *count, area* and *how.*

```
proc genmod data=ATPU.count_area_how;
      class area how;
     model count = area|how / dist = poisson link = log type3;
   output out = genout pred=fits resdev=res;
run;
```

Residual analysis

Since this is a saturated model, all residuals are zero.

Results

Analysis Of Parameter Estimates

Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept			1	0.6931	0.7071	-0.6928	2.0791	0.96
area	Avalon Peninsula		1	0.9163	0.8367	-0.7235	2.5561	1.20
area	Northeast Coast		1	-0.0000	1.0000	-1.9600	1.9600	0.00
area	Placentia Bay		1	-23.3863	1.1547	-25.6495	-21.1231	410.19
area	Straits		0	0.0000	0.0000	0.0000	0.0000	.
how	fishing gear		1	-0.6931	1.2247	-3.0936	1.7073	0.32
how	shot		0	0.0000	0.0000	0.0000	0.0000	.
area*how	Avalon Peninsula	fishing gear	1	2.6946	1.3142	0.1189	5.2703	4.20
area*how	Avalon Peninsula	shot	0	0.0000	0.0000	0.0000	0.0000	.
area*how	Northeast Coast	fishing gear	1	0.0000	1.7321	-3.3948	3.3948	0.00
area*how	Northeast Coast	shot	0	0.0000	0.0000	0.0000	0.0000	.
area*how	Placentia Bay	fishing gear	0	24.4849	0.0000	24.4849	24.4849	.
area*how	Placentia Bay	shot	0	0.0000	0.0000	0.0000	0.0000	.

```
   area*how   Straits           fishing gear   0     0.0000     0.0000     0.0000     0.0000       .
   area*how   Straits           shot           0     0.0000     0.0000     0.0000     0.0000       .
   Scale                                       0     1.0000     0.0000     1.0000     1.0000

                               Analysis Of Parameter Estimates

                       Parameter                                 Pr > ChiSq
                       Intercept                                     0.3270
                       area       Avalon Peninsula                   0.2734
                       area       Northeast Coast                    1.0000
                       area       Placentia Bay                      <.0001
                       area       Straits                             .
                       how        fishing gear                       0.5714
                       how        shot                                .
                       area*how   Avalon Peninsula  fishing gear     0.0403
                       area*how   Avalon Peninsula  shot              .
                       area*how   Northeast Coast   fishing gear     1.0000
                       area*how   Northeast Coast   shot              .
                       area*how   Placentia Bay     fishing gear      .
                       area*how   Placentia Bay     shot              .
                       area*how   Straits           fishing gear      .
                       area*how   Straits           shot              .
                       Scale

                              LR Statistics For Type 3 Analysis

                                                 Chi-
                            Source           DF     Square    Pr > ChiSq
                            area              3      28.63        <.0001
                            how               1       2.43        0.1191
                            area*how          3       9.23        0.0264
```

Declare decision

Reject $H_{o,}$ accept H_A that cause of death depends on area of recovery (G = 9.23, df=3, p = 0.0264). This p-value is within a factor of two of α, n is small and an analysis of residuals was not possible which makes me uneasy. If I were to publish this, I would consider using randomization or Fishers Exact Test.

Analysis of parameters

Since the interaction term was significant, no analysis of parameters for the main effects is possible. In order to compare netting with shooting in each geographic area, I reanalyzed each area separately using the "by" statement in PROC GENMOD.

```
proc genmod data=ATPU.how_where;
      by area;
      class  how ;
      model count = how / dist = poisson link = log type3;
run;
```

The following results were obtained (SAS output listing not shown):

Area	**Result**
Avalon Peninsula	$G = 27.56$, $df = 1$, $p < 0.0001$
Placentia Bay	Failed to converge due to 0 count
Straits	$G = 0.34$, $df = 1$, $p = 0.5599$
Northeast Coast	$G = 0.34$, $df = 1$, $p = 0.5599$

Thus, frequency of netting and shooting differ on the Avalon Peninsula ($G = 27.56$, $df = 1$, $p < 0.0001$), where death by fishing gear was 7.4 (i.e. $e^{2.0015}$) times more frequent than shooting (95% CI: 2.9 – 18.83).

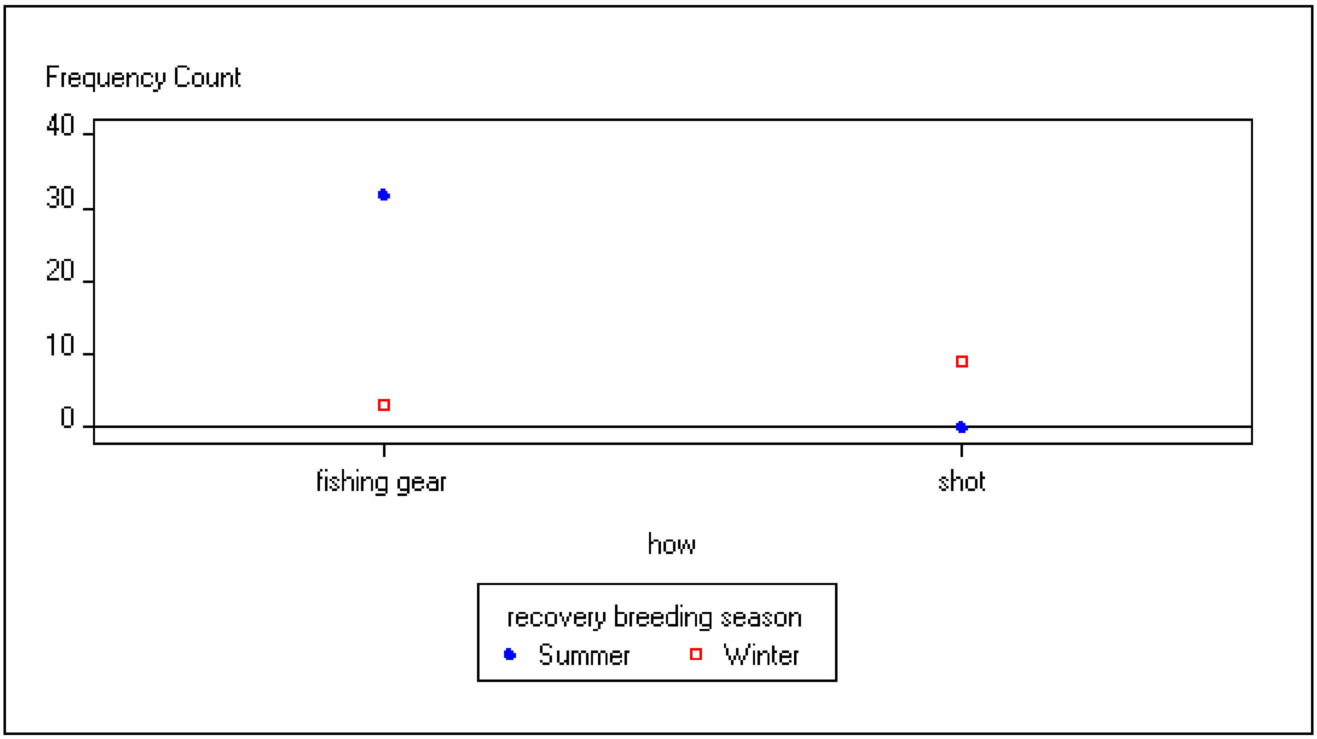
Recovery Frequency: Effect of Season and Cause of Death

Given the previous analysis I want to add the effect of season i.e. does seasonal recovery frequency depend on cause of death.

For this analysis I am interested in the two most common causes of death: fishing gear and shooting.

Verbal Model: recovery frequency depends upon the causes of death and season of recovery.

Graphical model:



Contingency Table

how	Summer	Winter
fishing gear	32	3
shot	0	9

Variables

Response: *count* = recovery frequency

Explanatory:

how = cause of death, on a nominal scale with two levels

s*eason* = time of year, on a nominal scale with two levels

Formal model:

$$\text{count} = e^{u} + \varepsilon$$

$$u = \beta_o + \beta_{how} \cdot \text{how} + \beta_{season} \cdot \text{season} + \beta_{how \cdot season} \cdot \text{how} \cdot \text{season}$$

Hypothesis

H$_A$: $\beta_{how \cdot season} \neq 0$ seasonal recovery frequency depends on the cause of death
H$_o$: $\beta_{how \cdot season} = 0$

Execution

Data in three columns *count*, *season* and *how*.

```
proc genmod data=ATPU.how_season;
      class how season;
     model count = how|season/ dist = poisson link = log type3;
   output out = genout pred=fits resdev=res;
run;
```

Residual analysis

Since this is a saturated model, all residuals are zero.

Results

Analysis Of Parameter Estimates

Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits	
Intercept			1	2.1972	0.3333	1.5439	2.8505
how	fishing gear		1	-1.0986	0.6667	-2.4053	0.2080
how	shot		0	0.0000	0.0000	0.0000	0.0000
season	Summer		1	-24.8904	0.6038	-26.0738	-23.7069
season	Winter		0	0.0000	0.0000	0.0000	0.0000
how*season	fishing gear	Summer	0	27.2575	0.0000	27.2575	27.2575
how*season	fishing gear	Winter	0	0.0000	0.0000	0.0000	0.0000
how*season	shot	Summer	0	0.0000	0.0000	0.0000	0.0000
how*season	shot	Winter	0	0.0000	0.0000	0.0000	0.0000
Scale			0	1.0000	0.0000	1.0000	1.0000

Analysis Of Parameter Estimates

Parameter			Chi-Square	Pr > ChiSq
Intercept			43.45	<.0001
how	fishing gear		2.72	0.0994
how	shot		.	.
season	Summer		1699.28	<.0001
season	Winter		.	.
how*season	fishing gear	Summer	.	.
how*season	fishing gear	Winter	.	.
how*season	shot	Summer	.	.
how*season	shot	Winter	.	.

LR Statistics For Type 3 Analysis

Chi-

```
                 Source              DF      Square      Pr > ChiSq
                 how                  1        8.41          0.0037
                 season               1        1.43          0.2320
                 how*season           1       31.09          <.0001
```

Declare decision

Reject H_o, accept H_A that seasonal recovery frequency depends on cause of death ($G = 31.09$, df=1, $p < 0.0001$).

Analysis of parameters
Since the interaction term is significant we need to fit a model for each season separately in order to analyze parameters of interest, as follows.

Formal Models:

For each of summer and winter:

$$\text{count} = e^u + \varepsilon$$

$$u = \beta_o + \beta_{how} \cdot \text{how}$$

Hypotheses

For winter:
H_A: $\beta_{how} \neq 0$ frequency of recoveries during winter depends on cause of death
H_o: $\beta_{how} = 0$

For summer:
H_A: $\beta_{how} \neq 0$ frequency of recoveries during summer depends on cause of death
H_o: $\beta_{how} = 0$

Execution
Use the PROC GENMOD "by" option to perform separate analyses for summer and winter:

```
Proc genmod data=ATPU.how_season;
      by season;
      class how;
      model count = how/ dist = poisson link = log type3;
run;
```

Residual analysis
Since these are saturated models, all residuals are zero.

Results

```
----------------------- season=Summer-----------------------

                        Analysis Of Parameter Estimates

                                         Standard       Wald 95%         Chi-
  Parameter                  DF   Estimate      Error    Confidence Limits    Square
```

```
Intercept                  1  -22.6932    0.1768  -23.0396  -22.3467  16479.3
how        fishing gear    0   26.1589    0.0000   26.1589   26.1589        .
how        shot            0    0.0000    0.0000    0.0000    0.0000        .
Scale                      0    1.0000    0.0000    1.0000    1.0000

                      Analysis Of Parameter Estimates

                    Parameter                Pr > ChiSq
                    Intercept                    <.0001
                    how        fishing gear       .
                    how        shot               .
                    Scale
                     LR Statistics For Type 3 Analysis

                                          Chi-
                Source           DF     Square    Pr > ChiSq
                how               1      44.36        <.0001

----------------------- season=Winter -----------------------
                                       Standard       Wald 95%          Chi-
Parameter                 DF  Estimate    Error   Confidence Limits   Square

Intercept                  1    2.1972    0.3333    1.5439    2.8505    43.45
how        fishing gear    1   -1.0986    0.6667   -2.4053    0.2080     2.72
how        shot            0    0.0000    0.0000    0.0000    0.0000        .
Scale                      0    1.0000    0.0000    1.0000    1.0000


                      Analysis Of Parameter Estimates

                    Parameter                Pr > ChiSq

                    Intercept                    <.0001
                    how        fishing gear      0.0994
                    how        shot               .
                    Scale

                     LR Statistics For Type 3 Analysis

                                          Chi-
                Source           DF     Square    Pr > ChiSq
                how               1       3.14        0.0764
```

Report Decision:

Summer:

When summer is analyzed separately, frequency of recovery depends on cause of death (G = 44.36, df=1, p < 0.0001). This should come as no surprise since there are no (reported) shooting deaths during summer and thus the parameter estimate comparing the two is non-sensible ($e^{26.1589}$).

Winter:

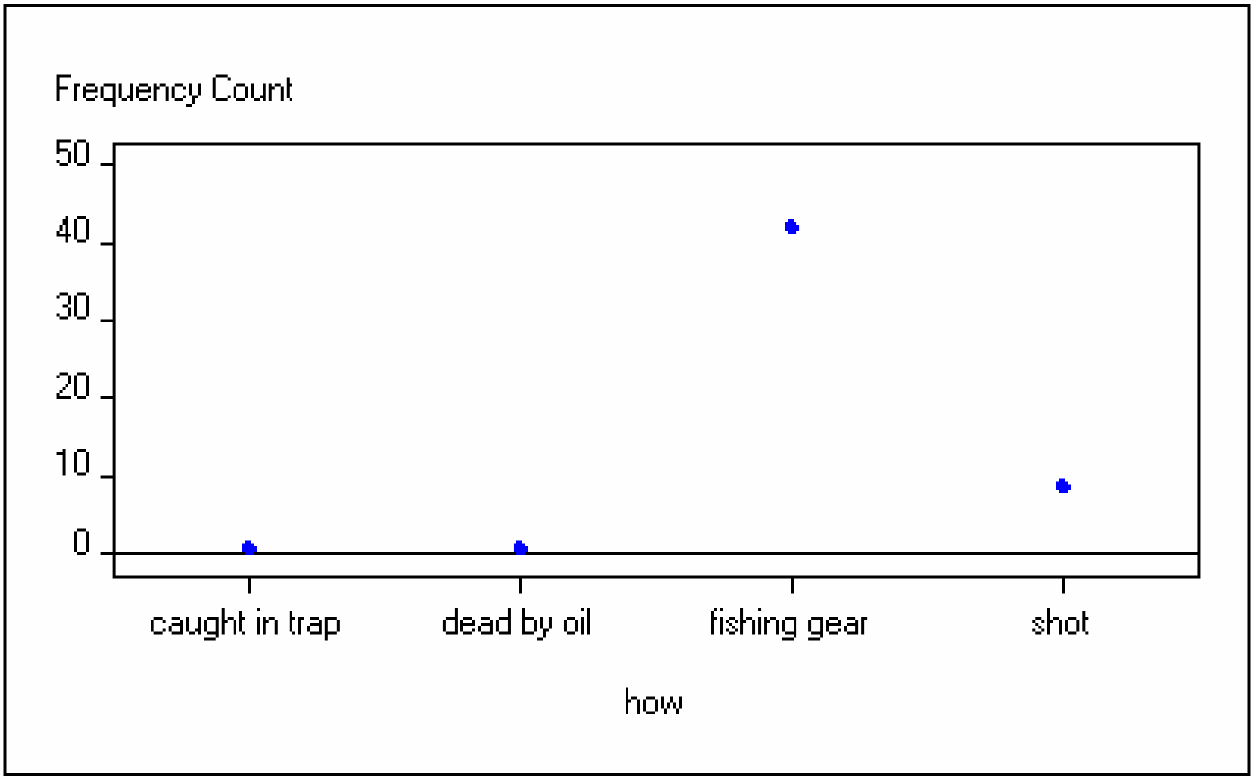
When winter is analyzed separately, frequency of recovery does not depend on type of death (G = 3.14, df=1, p = 0.0765).

Recovery Frequencies: Effect of Cause of Death

In this analysis I want to investigate the effect of cause of death on band recovery frequency.

Verbal Model: recovery frequency depends upon the method of recovery (i.e. how the bird died)

Graphical model:



Data Table

How	Count
caught in trap	1
dead by oil	1
fishing gear	42
shot	9

Variables

Response: *count* = frequency of recoveries, on a ratio scale

Explanatory: *how* = cause of death, on a nominal scale with four levels (caught in trap, dead by oil, fishing gear, shot)

Formal model:

count = $e^{u} + \varepsilon$

$u = \beta_o + \beta_{how} \cdot \text{how}$

Hypothesis

H_A: $\beta_{how} \neq 0$ frequency depends on how bird died
H_o: $\beta_{how} = 0$

Execution
Data in two columns: *count* and *how*.

```
proc genmod data=ATPU.how;
      class  how;
      model count = how / dist = poisson link = log type3;
run;
```

Residual analysis
Since this is a saturated model, all residuals are zero.

Results

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	2.1972	0.3333	1.5439	2.8505	43.45	<.0001
how	caught in trap	1	-2.1972	1.0541	-4.2632	-0.1312	4.35	0.0371
how	dead by oil	1	-2.1972	1.0541	-4.2632	-0.1312	4.35	0.0371
how	fishing gear	1	1.5404	0.3673	0.8205	2.2604	17.59	<.0001
how	shot	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
how	3	79.61	<.0001

Declare decision
Reject H_o, accept H_A that band recovery frequency depends on cause of death ($G = 79.61$, df=3, $p < 0.0001$).

Analysis of parameters
The two most frequent causes of death are entanglement in fishing gear and being shot. Entanglement in fishing gear is 4.67 ($e^{1.5405}$) times (95% CI: 2.27 - 9.58) more frequent than being shot.

Discussion

A summary of results is presented in Table 3 followed by a brief discussion.

Table 3: Summary of Results

Pg.	Result	Evidence
Error! Bookmark not defined.	Mean distance moved depends on age: adults: 88.4 (SE= 21.9) km juveniles: 654 (SE = 239) km	$F_{1, 65}$= 26.14, $p < 0.0001$
8	Mean distance moved depends on season: summer: 39.81 (SE = 7.91) km winter: 271.4 (SE = 89.3) km	$F_{1, 55}$= 19.94, $p < 0.0001$
11	Mean distance moved depends on cause of death: fishing net: 66.7 (SE = 17.4) km shot: 318 (SE = 116) km	$F_{1,48}$ = 12.71, $p < 0.0001$
14	Seasonal variation in distance moved is not significant when variation due to of cause of death is taken into account.	season: $F_{1, 40}$ = 0.05, p = 0.823 cause of death: $F_{1, 40}$ = 7.09, p = 0.011
16	Mean distance moved does not depend on year	$F_{1, 66}$ = 2.19, $p < 0.144$
20	Frequency of band returns decreasing by 7.05 %/year (95% CI: 3.19% - 10.76%)	G = 13.08, df = 1, p = 0.0003
28	Netting frequency declining 7.23%/year	G = 22.16, df = 1, $p < 0.0001$
31	Pre-moratorium netting frequency significantly different than post-moratorium: Mean pre-moratorium frequency: 1.67 (SE = 0.374) birds/yr Mean pre-moratorium frequency: 0.154 (SE = 0.104) birds/yr	G = 22.73, df = 1, $p < 0.0001$
34	No trends in netting mortalities when pre- and post-moratorium periods analyzed separately Pre-moratorium Post-moratorium	 G = 2.69, df = 1, p = 0.1010 G = 43.11, df = 1, p = 0.5091
39	Adults 13.24 (95% CI: 4.16 - 42.13) times more likely to be found than juveniles	G = 43.11, df = 1, $p < 0.0001$
41	Frequency of recovery in a given geographic area depends on season: Avalon Peninsula: 4.7 (95% CI: 2.27 - 9.59) times greater in summer than winter Northeast coast: not enough data	G = 9.99, df = 3, p = 0.0187 G = 23.17, df = 1, $p < 0.0001$

	Placentia Bay: No statistical difference in recovery frequency between summer and winter Straits: No statistical difference in recovery frequency between summer and winter	G = 0.00, df = 1, p = 1.0 G = 0.34, df = 1, p = 0.5599
46	Cause of death: frequency of netting versus shooting depends on geographical area Netting versus shooting by area: Avalon Peninsula: netting 7.4 (95% CI: 2.90 - 18.83) times more frequent than shooting Placentia Bay: not enough data Straits: no significant difference between netting and shooting Northeast coast: no significant difference between netting and shooting	G = 9.23, df = 3, p = 0.0264 G = 27.56, df = 1, p < 0.0001 G = 0.34, df = 1, p = 0.5599 G = 0.34, df = 1, p = 0.5599
50	Recovery frequency for a given cause of death depends on time of year Netting versus shooting by season: Summer: netting more frequent than shooting (no summer shooting records) Winter: no significant difference in recovery frequency for netting versus shooting	G = 31.09, df = 1, p < 0.0001 G = 44.36, df = 1, p < 0.0001 G = 3.14, df = 1, p = 0.0765
54	Entrapment in fishing gear 4.67 (95% CI: 2.27 - 9.58) times more frequent than shooting	G = 79.61, df = 3, p < 0.0001

There was an overall decrease of 7.05%/year in puffin recovery frequency in Newfoundland for the period 1968-2001 but no significant change in distances moved. Netting recovery rates have shown a significant decreasing trend during this period of this study. Changes in fishing practices since the 1992 Groundfish Moratorium appear to be driving this since mean netting recovery rates were significantly higher during pre- moratorium years compared to post- moratorium years and there was no significant linear trend in either of these two periods when analyzed separately.

Adult birds were recovered more frequently and closer to the breeding colony than juveniles. Juveniles are not tied to the breeding colony for at least the first three years of their life so they can wander further to places where they are unlikely to be recovered when they die.

In Newfoundland, recovery frequency varies by geographic region, season and cause of death. Puffins face a variety of threats including oiling, entrapment in fishing gear (netting) and accidental shooting in the Newfoundland murre hunt. Netting and shooting are the two most frequent causes of death reported; oiled birds sink at sea and are likely underreported in this data set. Entanglement in fishing gear happens closer to the colony and more in summer while shooting occurs further from the colony in winter. Recoveries are more frequent on the Avalon Peninsula than elsewhere, with more recoveries there in summer than in winter and more death by fishing gear there than by other means.

The band-return patterns explored herein are intimately linked with patterns of human population density and behavior and this bias must always be kept in mind when drawing conclusions from band-return data. This may be particularly true for seabirds that spend much of their life on the open ocean largely inaccessible to humans. In some cases, it may be that band-return patterns are a better indication of human behavior than they are of bird behavior.

Reference List

Agresti, A 1996. Introduction to Categorical Data Analysis. John Wiley & Sons, NY.

Brown, R. G. B. 1985. The Atlantic Alcidae at Sea. *In* The Atlantic Alcidae: evolution, distribution and biology of the auks inhabiting the Atlantic Ocean and adjacent water areas. (D. N. Nettleship and T. R. Birkhead, eds.). Academic Press, London. pp. 383-426.

Brown, R G B 1986. Revised Atlas of Eastern Canadian Seabirds. Canadian Wildlife Service, Ottawa, ON.

Croxall, J P 1984. Status and conservation of the world seabirds. International Council for Bird Preservation, Cambridge, England.

Dennis, J. V. 1981. A summary of banded North American birds encountered in Europe. North American Bird Bander **6:** 88-96.

Harris, M. P. 1984. Movements and mortality patterns of north Atlantic Puffins as shown by ringing. Bird Study **31:** 131-140.

Harrison, P 1983. Seabirds: an identification guide. Houghton Mifflin Co., Boston, MA.

Littell, R C, Stroup, W W, and Freund, R J 2002. SAS for Linear Models. SAS Institute Inc., Cary, NC.

Lock, A R, Brown, R G B, and Gerriets, S H 1994. Gazetteer of Marine Birds in Atlantic Canada: An Atlas of Vulnerability to Oil Pollution. Canadian Wildlife Service, Atlantic Region.

Lowther, P. E., Diamond, A. W., Kress, S. W., Robertson, G. J., and Russell, K. 2002 Atlantic Puffin (*Fratercula arctica)*. Poole, A. and Gill, F. *In* The Birds of North America no. 709. The Birds of North America, Inc., Philadelphia, PA.

Mead, C. J. 1974. The results of ringing Auks in Britain and Ireland. Bird Study **21:** 45-86.

Nettleship, D. N. and Evans, P. G. H. 1985. Distribution and status of the Atlantic Alcidae. *In* The Atlantic Alcidae: evolution, distribution and biology of the auks inhabiting the Atlantic Ocean and adjacent water areas. (D. N. Nettleship and T. R. Birkhead, eds.). Academic Press, London, UK. pp. 54-154.

Stokes, M E, Davis, C S, and Koch, G G 2000. Categorical Data Analysis Using the SAS System. SAS Institute Inc., Cary, NC.