

Lab 5: Goodness-of-fit Testing in Program Mark

Objectives: The purpose of this exercise is to introduce you to the procedures used for Goodness-of-fit testing in Program Mark. GOF testing allows you to determine whether the assumptions of mark-recapture analyses are met by the data in the capture histories. It also allows you to correct for any lack of fit between your global starting model and your data by calculating the variance inflation factor, $c\text{-hat}$. This parameter can then be used to adjust AICc values to calculate QAICc instead. We will apply three GOF procedures to the Snowy Plover data set.

Locate the database file *snowy.dbf* on the computer from last week and open it in Program Mark. Be sure that the standard CJS model $\phi(g^*t), p(g^*t)$ is among the candidate models.

1. Program Release GOF - A GOF Test to $\phi(g^*t), p(g^*t)$ with diagnostics to sort out problems.

Click on Tests | Program Release GOF

The results should open in a Wordpad window. Nothing too complicated in implementing this analysis, all of the results are in this window. The trick is figuring out what all of this means.

The top lines go through some of the changes to the coding of Release made years ago. It then regurgitates the capture histories:

```
INPUT --- proc chmatrix occasions=4 groups=2 FPool;

INPUT ---      glabel(1)=Males;

INPUT ---      glabel(2)=Females;
INPUT ---      1000  9 18;
INPUT ---      1001  2  1;
INPUT ---      1010  1  8;
INPUT ---      1011  0  4;
INPUT ---      1100  6 12;
INPUT ---      1101  3  2;
INPUT ---      1110  6  2;
INPUT ---      1111  4  4;
INPUT ---      0100 37 48;
INPUT ---      0101  4  7;
INPUT ---      0110 15 11;
INPUT ---      0111 12  8;
INPUT ---      0010 31 44;
INPUT ---      0011 32 30;
```

and then calculates the reduced m-array for each group. In this table, $R(i)$ is the number released at each occasion (including individuals marked on previous occasions), $m(i,j)$ is the number recaptured on subsequent occasions, and $r(i)$ is the total number recaptured. If the probability of encounter is high ($p > 0.7$) then most of the recaptures will be along the diagonal. If the probability of encounter is low ($p < 0.7$), then most of the recaptures will be along the off-diagonal. Note that data can be imported into Program Mark in the format of a reduced m-array.

Observed Recaptures for Group 1

Males					
i	R(i)	m(i, j)			r(i)
		j=	2	3	4
1	31	19	1	2	22
2	87		37	7	44
3	101			48	48
m(j)		19	38	57	
z(j)		3	9	0	

Observed Recaptures for Group 2
Females

i	R(i)	m(i, j)			r(i)
		j=	2	3	4
1	51	20	12	1	33
2	94		25	9	34
3	111			46	46
m(j)		20	37	56	
z(j)		13	10	0	

Sums for the above Groups

m.	0	39	75	113
z.	0	16	19	0
R.	82	181	212	
r.	55	78	94	

The next steps are the components of TESTS 3 and 2 for each transition. Refer to the lecture handout to remember what the labels for rows and columns are. O = observed count, E = expected count, and C = the χ^2 value for that cell. Fisher's exact test is a different contingency test for 2 x 2 tables.

Test for Group 1
Males

TEST 3.SR2: Animals captured on occasion 2

+-----+-----+					
O	13		6		19
E	9.6		9.4		
C	1.2		1.2		
+-----+-----+					
O	31		37		68
E	34.4		33.6		
C	0.3		0.3		
+-----+-----+					
	44		43		87

Chi-square=3.0973 (df=1) P=0.0784
Fisher's Exact Test P=0.1187

TEST 3.SR3: Animals captured on occasion 3

O	16	22	38
E	18.1	19.9	
C	0.2	0.2	
O	32	31	63
E	29.9	33.1	
C	0.1	0.1	

48 53 101
 Chi-square=0.7175 (df=1) P=0.3970
 Fisher's Exact Test P=0.4187

Goodness of fit test of seen before versus not seen before against when next seen again by capture occasions.

Test for Group 1
 Males

TEST 3.Sm2: Animals captured on occasion 2

O	27	4	31
E	26.1	4.9	
C	0.0	0.2	
O	10	3	13
E	10.9	2.1	
C	0.1	0.4	

37 7 44
 Chi-square=0.7086 (df=1) P=0.3999
 Fisher's Exact Test P=0.6542

Test for Group 2
 Females

TEST 3.SR2: Animals captured on occasion 2

O	8	12	20
E	7.2	12.8	
C	0.1	0.0	
O	26	48	74
E	26.8	47.2	
C	0.0	0.0	

34 60 94
 Chi-square=0.1614 (df=1) P=0.6879
 Fisher's Exact Test P=0.7943

TEST 3.SR3: Animals captured on occasion 3

O	16	21	37
E	15.3	21.7	
C	0.0	0.0	
O	30	44	74
E	30.7	43.3	
C	0.0	0.0	

46 65 111
 Chi-square=0.0742 (df=1) P=0.7852
 Fisher's Exact Test P=0.8395

Test for Group 2
 Females

TEST 3.Sm2: Animals captured on occasion 2

O	19	7	26
E	19.1	6.9	
C	0.0	0.0	
O	6	2	8
E	5.9	2.1	
C	0.0	0.0	

25 9 34
 Chi-square=0.0116 (df=1) P=0.9141
 Fisher's Exact Test P=1.0000

Cumulative result of TEST 3 over groups
 Chi-square=3.3974 (df=6) P=0.7576

Goodness of fit test of recaptures partitioned by rows.

For reason that is unclear to be, the version of Program Release bundled with Mark does not calculate both of the component tests of TEST 2. What is labeled as '2.Cn' is the same as Test 2.Ctn.

Test for Group 1
 Males

TEST 2.C2: Test of row 1 vs. row 2

O	1	2	3
E	2.4	0.6	
C	0.8	3.5	
O	37	7	44
E	35.6	8.4	
C	0.1	0.2	

38 9 47
 Chi-square=4.6736 (df=1) P=0.0306
 Fisher's Exact Test P=0.0895

Test for Group 2
Females

TEST 2.C2: Test of row 1 vs. row 2

+-----+-----+			
O	12	1	13
E	10.2	2.8	
C	0.3	1.1	
+-----+-----+			
O	25	9	34
E	26.8	7.2	
C	0.1	0.4	
+-----+-----+			
	37	10	47

Chi-square=1.9798 (df=1) P=0.1594
Fisher's Exact Test P=0.2437

Cumulative result of TEST 2 over groups
Chi-square=4.2405 (df=2) P=0.1200

The final output is a summary table that pools across the various components. All of the χ^2 tests are additive and can be pooled for an overall GOF test. Early versions of Release report just the χ^2 values and their *P*-values. The version of Program Release currently bundled with Mark reports the *P*-values associated with the Fisher Exact test and then works backwards to generate the χ^2 values associated with that *P*-value and *df*. This is apparently a robust procedure. If you discover discrepancies between the test statistics of different version of Release, be aware that this is the likely explanation.

Summary of TEST 3 (Goodness of fit) Results

Group	Component	X ²	df	P-level	Sufficient Data
1	3.SR2	2.4339	1	0.1187	Yes
1	3.SR3	0.6538	1	0.4187	Yes
1	3.Sm2	0.2006	1	0.6542	Yes
Group 1	TEST 3	3.2884	3	0.3493	
2	3.SR2	0.0680	1	0.7943	Yes
2	3.SR3	0.0411	1	0.8395	Yes
2	3.Sm2	0.0000	1	1.0000	Yes
Group 2	TEST 3	0.1090	3	0.9907	
All Groups	TEST 3	3.3974	6	0.7576	

Summary of TEST 2 (Goodness of fit) Results

Group	Component	X ²	df	P-level	Sufficient Data
1	2.C2	2.8813	1	0.0895	No
2	2.C2	1.3592	1	0.2437	Yes
All Groups	TEST 2	4.2405	2	0.1200	

Goodness of Fit Results (TEST 2 + TEST 3) by Group

Group	X ²	df	P-level
1	6.1697	4	0.1868
2	1.4682	4	0.8323
Total	7.6379	8	0.4696

The overall GOF test is nonsignificant, indicating that the data are a good fit to the CJS model $\phi_{(g^{*t})}, p_{(g^{*t})}$. If we wanted to calculate a variance inflation factor it would be:

$$\hat{c} = \frac{\sum(\text{Test 2 } \chi^2 + \text{Test 3 } \chi^2)}{\sum(\text{Test 2 df} + \text{Test 3 df})}$$

$$\hat{c} = 7.64 / 8 = 0.96$$

Variance inflation factors are asymptotically expected to be 1 if there is no overdispersion due to pseudoreplication or heterogeneity in the capture histories. Thus, \hat{c} -values ≤ 1 are set to 1.

The remainder of the output of Program Release below the above table is fitting various canned models that are better addressed with Program Mark.

2. Bootstrap GOF - A parametric bootstrap test applied to the starting global model.

Find the CJS model in the set of candidate models $\phi(g^{*t}), p(g^{*t})$ and highlight it. Click Retrieve | Current Model to load this model. Write down the deviance for this model which will be the observed deviance.

Click on Tests | Bootstrap GOF

Select Deviance Only (fastest) and click OK.

This gives a dialog box where you can specify the number of bootstrap replicates to run. They are often a bit slow so accept the default value of 100 replicates. In practice, you would likely run 1000 to 10,000 replicates. Click OK to start the simulation. Mark will prompt you for the name of a new file to put the bootstrap results in. The default is BootstrapResults.dbf, you might want to label the output file SnowyBoot.dbf instead. Click OK and sit back and let the simulation run to completion.

A simulation progress box monitors the progress of the Bootstrap GOF test in Program Mark. It is not unusual for these simulation to jam and give you cryptic error messages. If successful, this procedure will generate a distribution of deviance values by bootstrapping your capture histories as if they met the assumptions of independence and homogeneity.

When the simulations are complete you will see your usual Results window and no bootstrap output. Close the Results window and click

File | Simulations | View Simulation File. Select and open SnowyBoot.dbf. This file gives a list of all of the bootstrap replicates.

Click the calculator on the right to get the basic stats for the expected deviance of the bootstrap distribution based on 100 replicates (mean, SE, 95%CI). You can then calculate the variance inflation factor by hand as:

$$\hat{c} = \text{observed deviance} / \text{mean expected deviance}$$

To obtain the observed deviance, you may need to return to your model Results window and look up the deviance value for the global model $\phi(g^*t)$, $p(g^*t)$

Ideally the \hat{c} -value is close to 1 if the model assumptions are met.

To get a P -value for the null hypothesis that $\hat{c} = 1$, go back to the bootstrap output. You will need to look at where the observed deviance falls in the distribution of the expected values. Close the Wordpad view with the mean expected deviance and return to the bootstrap file. Click the *A-Z arrow* button, Select Deviance and click OK. These steps will sort all of the expected deviances in the bootstrap distribution from low to high. Determine the rank of the observed deviance by scanning through the file to find the two consecutive replicates that straddle the observed value. Highlight the line immediately above where the observed value would fall and click the *blue arrow* button to obtain the rank. If the observed value falls between expected values with rank 62 and rank 63, then the rank is 62.5. The P -value of the \hat{c} value can be calculated as: $1 - (\text{observed rank}/\text{total number of replicates})$. If the P -value is nonsignificant then the hypothesis of no overdispersion is accepted and you can proceed with model testing with confidence. If it is rejected but $\hat{c} < 3$ then you can simply account for the uncertainty. If \hat{c} is > 3 then you might want to consider a different starting model.

3. Median \hat{c} -hat procedure - a new approach to GOF testing

Again, find the CJS model in the set of candidate models $\phi(g^*t)$, $p(g^*t)$ and highlight it. Click Retrieve | Current Model to load this model.

Click on Tests | Median \hat{c} -hat

Program Mark sets bounds for estimated \hat{c} -hat values as: 1 for the lower bound, and the model deviance (Dev) divided by the number of parameters (K) for the upper bound ($13.966/10 = 1.40$). It then prompts you to set the number of intermediate points and the number of replicates per point. Accept the default values of 10 and 10 and click okay to run.

Like the Bootstrap GOF test, a simulation progress box monitors the progress of the Median \hat{c} -hat procedure in Program Mark. When completed Program Mark will generate an output window entitled 'Logistic Regression Estimate of \hat{c} '. The second line of this file will be the estimate of \hat{c} -hat produced by the Median \hat{c} -hat procedure. Program Mark also generates a graphics window which will show the results of this procedure including the logistic regression curve.

4. Making adjustments for $\hat{c} > 1$

Click Adjustments | \hat{c} and enter some value > 1 to apply the overdispersion factor to all of the models that you have tested for Snowy Plovers. Try $\hat{c} = 2$ and then $\hat{c} = 6$ to see what happens. For any value of $\hat{c} > 1$, the headers on the columns are immediately updated to QAICc from AICc to reflect that the adjustments for \hat{c} have been made. Note that the best fit model may change if you vary the \hat{c} value. In general, the higher the value of \hat{c} the more likely the model will collapse to something simple like $\phi(\cdot)$, $p(\cdot)$

The confidence limits around the parameter estimates also change. Click the button to the right of the garbage can ('View output from highlighted model...'), and scroll down to the bottom to see the SE for the real parameter estimates. Jot down a few of the SE. Close this window. Now click the third button to the right of the garbage can ('View estimates of real parameters...'). Inspect the SE for the same set of parameters. If \hat{c} is > 1 then the SE for the real estimates are larger to account for the uncertainty due to overdispersion. The adjusted value of \hat{c} should also appear somewhere in this output file.

Adjustments to \hat{c} can also be used to evaluate the sensitivity of mark-recapture analyses where no goodness-of-fit test is available. If no GOF test is available, two general approaches to dealing with overdispersion can be used. First, you could systematically vary \hat{c} within a range of possible overdispersion (e.g., \hat{c} values of 1 to 3 by 0.5) and look at how your model selection is affected. Alternatively, you could adjust \hat{c} by trial and error to determine the \hat{c} value that would be necessary to change the rankings of your models. If the best fit model when $\hat{c} = 1$ remains the best fit model until $\hat{c} > 7$, then your conclusions are clearly robust to the typical amounts of overdispersion found in count data ($\hat{c} = 1$ to 3).