

Exercices tests d'ajustement

Le but de cet exercice est de se familiariser avec les données de CR, d'ajuster par maximum de vraisemblance quelques modèles simples, de comparer ces modèles entre eux pour déterminer celui qui fournit la meilleure description des données et de tester la qualité de l'ajustement de ces modèles. Nous utiliserons les logiciels MARK et U-CARE (téléchargeables gratuitement depuis <http://www.cefe.cnrs.fr/BIOM/logiciels.htm>). Un corrigé sommaire est disponible sur l'EP dans le répertoire TP_CMV.

Question 1

On a simulé 2 jeux de données CR avec les paramètres de survie (ϕ) et recapture (p) suivants :

- jeu de données G1: $\phi=0.8$, $p=0.8$

- jeu de données G2: $\phi=0.8$, $p=0.2$.

A l'aide du logiciel MARK, ajuster séparément à G1 et G2 le modèle $\Phi(t) p(t)$ (appelé aussi modèle de Cormack-Jolly-Seber, ou CJS). Que pouvez-vous vous dire sur l'estimation des paramètres ?

Pour G1 :

Real Function Parameters of {Phi(t) p(t) Design Matrix}				
Parameter	Estimate	Standard Error	95% Confidence Interval Lower	Upper
1:Phi	0.7955353	0.0105277	0.7741270	0.8153983
2:Phi	0.7976064	0.0104554	0.7763388	0.8173274
3:Phi	0.7972201	0.0104539	0.7759580	0.8169405
4:Phi	0.7974462	0.0104689	0.7761511	0.8171926
5:Phi	0.7971865	0.0105930	0.7756312	0.8171591
6:Phi	0.7965998	0.0114273	0.7732835	0.8180823
7:Phi	0.7579586	0.0000000	0.7579586	0.7579586
8:p	0.8057625	0.0113500	0.7825465	0.8270477
9:p	0.8047457	0.0105751	0.7831830	0.8246441
10:p	0.8047680	0.0104631	0.7834424	0.8244641
11:p	0.8053606	0.0104541	0.7840502	0.8250373
12:p	0.8047207	0.0105763	0.7831555	0.8246215
13:p	0.8057626	0.0113500	0.7825466	0.8270478
14:p	0.8414490	0.0000000	0.8414490	0.8414490

Pour G2 :

g2				
Real Function Parameters of {Phi(t) p(t) Design Matrix}				
Parameter	Estimate	Standard Error	95% Confidence Interval Lower	95% Confidence Interval Upper
1:Phi	0.7881391	0.0284981	0.7269548	0.8386560
2:Phi	0.8034371	0.0289777	0.7404441	0.8541533
3:Phi	0.8011612	0.0302072	0.7353440	0.8538633
4:Phi	0.7982039	0.0324221	0.7272242	0.8544114
5:Phi	0.7998245	0.0370273	0.7174630	0.8627694
6:Phi	0.8011440	0.0483595	0.6896532	0.8795759
7:Phi	0.3799300	6.5176952	0.1727051E-023	1.0000000
8:p	0.1997322	0.0122095	0.1768720	0.2247403
9:p	0.2000415	0.0099724	0.1812090	0.2203046
10:p	0.1993093	0.0093410	0.1816301	0.2182503
11:p	0.1996119	0.0093509	0.1819133	0.2185723
12:p	0.1996131	0.0099628	0.1808002	0.2198582
13:p	0.1991477	0.0121132	0.1764636	0.2239548
14:p	0.4178089	7.1675090	0.5939178E-025	1.0000000

⇒ Les estimations sont très satisfaisantes, sauf pour la probabilité de survie sur le dernier intervalle de temps et la probabilité de recapture à la dernière occasion d'échantillonnage. Ce phénomène est connu et vient du fait que ces 2 paramètres ne sont pas estimables séparément mais seulement sous la forme de leur produit : on parle de redondance en paramètres. C'est un phénomène important à prendre en compte pour l'interprétation des paramètres bien évidemment, mais aussi pour le calcul de l'AIC pour la sélection de modèle, puisque c'est le nombre de paramètres estimables qui rentre en jeu dans ce calcul. Ici pour le modèle CJS, il y a donc 1 paramètre estimable de moins que le nombre total de paramètres dans le modèle.

Question 2

a) Grouper les jeux de données G1 et G2 pour obtenir le jeu de données G1+G2.

Il suffit de mettre les 2 fichiers de données bout à bout dans un seul et même nouveau fichier (attention, ce fichier doit avoir l'extension .inp pour être reconnu par le logiciel).

b) Ajuster le modèle CJS à G1+G2. Que remarquez-vous concernant l'estimation des paramètres ?

G1+G2				
Real Function Parameters of {Phi(t) p(t) Design Matrix}				
Parameter	Estimate	Standard Error	95% Confidence Interval Lower	95% Confidence Interval Upper
1:Phi	0.7221845	0.0112559	0.6995938	0.7436987
2:Phi	0.7395884	0.0115847	0.7162503	0.7616458
3:Phi	0.7408966	0.0119107	0.7168775	0.7635501
4:Phi	0.7333956	0.0122022	0.7088080	0.7566200
5:Phi	0.7162467	0.0125988	0.6909187	0.7402793
6:Phi	0.6861635	0.0137396	0.6586288	0.7124470
7:Phi	0.6400092	0.0000000	0.6400092	0.6400092
8:p	0.5521954	0.0114437	0.5296746	0.5745041
9:p	0.5114639	0.0099805	0.4918945	0.5309983
10:p	0.4942521	0.0095609	0.4755296	0.5129908
11:p	0.4947520	0.0095645	0.4760218	0.5134970
12:p	0.5112464	0.0099823	0.4916739	0.5307846
13:p	0.5502705	0.0114173	0.5278058	0.5725320
14:p	0.6211459	0.0000000	0.6211459	0.6211459

→ phi est affecté (plus variable et entre 0.65 et 0.75), et p est estimé à une valeur intermédiaire autour de 0.5

Question 3

A l'aide du logiciel U-CARE, tester la qualité de l'ajustement du modèle CJS aux données

G1, G2 et G1+G2. Quelles sont vos conclusions ?

Les tests d'ajustement servent à détecter les écarts aux hypothèses

Tests d'ajustement sur G1 : OK

- Global TEST, number of groups =1
- df =24
- Quadratic Chi2 =2.8188
- ->P-level=1
- N(0,1) statistic for transient(>0) =-0.30065
- ->P-level, two-sided test =0.76368
- ->P-level, one-sided test for transience =0.61816
- N(0,1) signed statistic for trap-dependence =1.0212
- ->P-level, two-sided test =0.30716

Tests d'ajustement sur G2 : OK

- Global TEST, number of groups =1
- df =36
- Quadratic Chi2 =0.78875
- ->P-level=1
- N(0,1) statistic for transient(>0) =-0.60849
- ->P-level, two-sided test =0.54287
- ->P-level, one-sided test for transience =0.72857
- N(0,1) signed statistic for trap-dependence =-0.11582
- ->P-level, two-sided test =0.9078

Tests d'ajustement sur G1+G2 : significatif !

- Global TEST, number of groups =1
- df =36

- Quadratic Chi2 =1606.1382
- ->P-level=0
- N(0,1) statistic for transient(>0) =26.248 • ->P-level, two-sided test =0
- ->P-level, one-sided test for transience =0
- N(0,1) signed statistic for trap-dependence =-20.6059
- ->P-level, two-sided test =0

Question 4

Il peut y avoir des animaux en transit sur la zone d'étude.

- a) Pour créer artificiellement une telle situation, rajouter 100 individus en transit (i.e. possédant une histoire avec un seul événement de capture) à chaque date dans G1.

Voir fin du fichier G1transit.inp

- b) Ajuster le modèle CJS à ces nouvelles données. Quelles sont vos conclusions concernant les estimations ?

G1 transit				
Real Function Parameters of {Phi(t) p(t) Design Matrix}				
Parameter	Estimate	Standard Error	95% Confidence Interval Lower	Upper
1:Phi	0.7646056	0.0109328	0.7425071	0.7853558
2:Phi	0.7668174	0.0108571	0.7448664	0.7874189
3:Phi	0.7664675	0.0108530	0.7445266	0.7870629
4:Phi	0.7666400	0.0108674	0.7446683	0.7872612
5:Phi	0.7664370	0.0109859	0.7442194	0.7872764
6:Phi	0.7658146	0.0117860	0.7419293	0.7881205
7:Phi	0.9167131	14.088312	0.9399584E-156	1.0000000
8:p	0.7979545	0.0117187	0.7740098	0.8199522
9:p	0.7969108	0.0109213	0.7746638	0.8174788
10:p	0.7969300	0.0108065	0.7749256	0.8172904
11:p	0.7975443	0.0107976	0.7755549	0.8178853
12:p	0.7968811	0.0109228	0.7746313	0.8174520
13:p	0.7979545	0.0117187	0.7740099	0.8199523
14:p	0.6623602	10.179355	0.3530968E-038	1.0000000

→ phi sous-estimé (p OK)

- c) Tester l'ajustement du modèle CJS à ces mêmes données. Interpréter en particulier la composante 3.SR du test.

- Global TEST, number of groups =1
- df =24
- Quadratic Chi2 =175.9949
- ->P-level=0
- N(0,1) statistic for transient(>0) =13.0784
- ->P-level, two-sided test =0

- ->P-level, one-sided test for transience =0
- $N(0,1)$ signed statistic for trap-dependence =1.0212
- ->P-level, two-sided test =0.30716

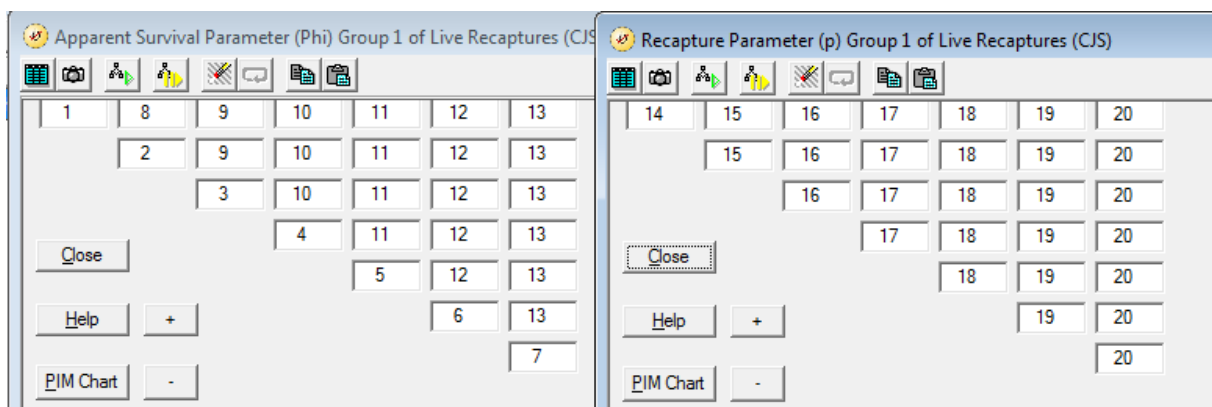
Test d'ajustement : une composante est très significative (3.SR). Dans ce test, on teste l'hypothèse nulle suivante pour chaque occasion de capture : « Il n'y a pas de différence dans la probabilité d'être revu après l'occasion i entre les nouveaux (à l'occasion i) et les anciens individus ». Cette hypothèse est rejetée : aussi le modèle CJS n'est pas adapté à nos données.

Cependant, pour cette composante du test, nous pouvons prendre en compte cette différence entre les nouveaux individus (dont on a introduit une partie) et les anciens. Si on y réfléchit, on a introduit des individus qui meurent aussitôt donc on attend une baisse de la survie juste après la première capture.

Il suffit pour cela dans le nouveau modèle de différencier les survies des nouveaux individus (survie : $\phi' = 0.8 \times \text{probabilité de rester}$), des autres (survie : $\phi = 0.8$). ϕ' sera la survie apparente (des individus restant) l'année après leur première capture (première classe d'âge dans la PIM). ϕ sera la survie des individus qui sont restés au moins un an (classes d'âge suivante dans la PIM).

d) Ajuster un modèle à 2 classes d'âge sur la survie $\phi(a2 \cdot t)$. Vos conclusions ?

voici la PIM nécessaire pour ajuster les survie ϕ et ϕ'



G1 transit				
Real Function Parameters of {Phi(a2*t) p(t)}				
Parameter	Estimate	Standard Error	95% Confidence Interval	
			Lower	Upper
1:Phi	0.7574128	0.0107943	0.7356346	0.7779385
2:Phi	0.7021461	0.0174685	0.6668175	0.7352161
3:Phi	0.6694859	0.0204818	0.6281848	0.7083294
4:Phi	0.6620311	0.0211602	0.6193987	0.7021841
5:Phi	0.6609527	0.0213794	0.6178781	0.7015175
6:Phi	0.6601383	0.0221386	0.6155033	0.7020996
7:Phi	0.6736438	0.0000000	0.6736438	0.6736438
8:Phi	0.7965637	0.0129874	0.7699221	0.8208383
9:Phi	0.7966663	0.0120975	0.7719250	0.8193518
10:Phi	0.7970065	0.0119845	0.7725033	0.8194872
11:Phi	0.7965886	0.0120991	0.7718444	0.8192774
12:Phi	0.7962885	0.0129834	0.7696577	0.8205582
13:Phi	0.8101903	0.0000000	0.8101903	0.8101903
14:p	0.8055323	0.0114590	0.7820858	0.8270141
15:p	0.8046143	0.0106464	0.7829013	0.8246420
16:p	0.8046544	0.0105285	0.7831908	0.8244694
17:p	0.8052223	0.0105206	0.7837720	0.8250198
18:p	0.8046143	0.0106464	0.7829014	0.8246420
19:p	0.8055322	0.0114590	0.7820857	0.8270140
20:p	0.7860444	0.0000000	0.7860444	0.7860444

Le paramètre phi' (age 1, soit les paramètres de 1 à 7) est biaisé, il représente le produit de la survie phi par la probabilité de rester, alors que phi (age 2, soit les paramètres de 8 à 13) est estimé correctement à 0.8.