

ROC解析



北海道大学 医学統計学
横田 勲

1

今回の内容

- ▶ 変数選択
- ▶ ROC解析

2

2

サンプルデータを利用

3

- ▶ ヘルプ> サンプルデータ
- ▶ 探索的モデリングのTitanic Passengers

3

データテーブル

4

The screenshot shows the JMP Pro interface with the Titanic Passengers data table loaded. The table has the following columns: 客室クラス (Cabin Class), 生存か否か (Survived), 氏名 (Name), 性別 (Sex), 年齢 (Age), 兄弟/配偶者の数 (Number of Siblings/Spouse), 親/子の数 (Number of Parents/Children), チケット番号 (Ticket Number), 料金 (Fare), and キャンビン (Cabin). The table contains 20 rows of data, with the first row being the header row. The data is as follows:

客室クラス	生存か否か	氏名	性別	年齢	兄弟/配偶者の数	親/子の数	チケット番号	料金	キャンビン
1	Yes	Allen, Miss. Elis...	female	29	0	0	24160	211.3375	B5
2	Yes	Allison, Master...	male	0.9167	1	2	113781	151.55	C2
3	No	Allison, Miss. H...	female	2	1	2	113781	151.55	C2
4	No	Allison, Mr. Hu...	male	30	1	2	113781	151.55	C2
5	No	Allison, Mrs. H...	female	25	1	2	113781	151.55	C2
6	Yes	Anderson, Mr. ...	male	48	0	0	19952	26.55	E1
7	Yes	Andrews, Miss...	female	63	1	0	13502	77.9583	D7
8	No	Andrews, Mr. T...	male	39	0	0	112050	0	A3
9	Yes	Appleton, Mrs. ...	female	53	2	0	11769	51.4792	C11
10	No	Artagaveytia, ...	male	71	0	0	PC 17609	49.5042	
11	No	Astor, Col. Joh...	male	47	1	0	PC 17757	227.525	C6
12	Yes	Astor, Mrs. Joh...	female	18	1	0	PC 17757	227.525	C6
13	Yes	Aubart, Mme. ...	female	24	0	0	PC 17477	69.3	B3
14	Yes	Barber, Miss. E...	female	26	0	0	19877	78.85	
15	Yes	Barkworth, Mr. ...	male	80	0	0	27042	30	A2
16	No	Baumann, Mr. ...	male		0	0	PC 17318	25.925	
17	No	Baxter, Mr. Qui...	male	24	0	1	PC 17558	247.5208	B5
18	Yes	Baxter, Mrs. Ja...	female	50	0	1	PC 17558	247.5208	B5
19	Yes	Bazzani, Miss. ...	female	32	0	0	11813	76.2917	D1
0									

4

変数選択

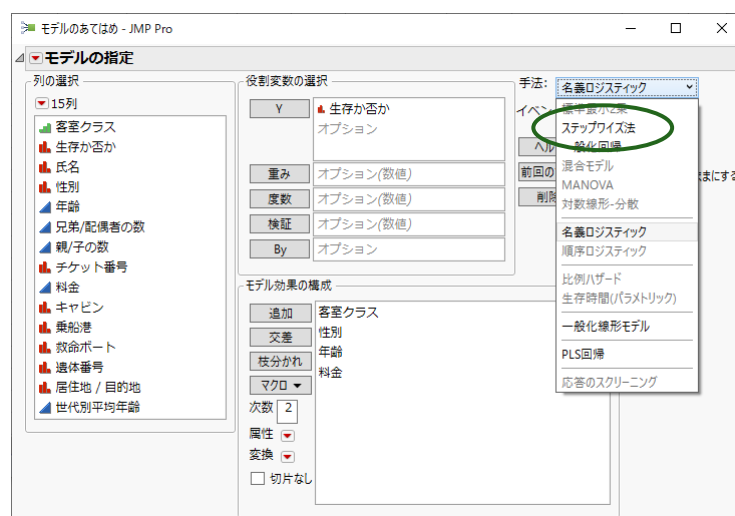
5

- ▶ 分析> モデルのあてはめ
- ▶ 生存か否かをYに
- ▶ 客室クラス、性別、年齢、料金をモデル効果の構成に

5

手法をステップワイズに

6



6

ステップワイズ

7

(-1)*対数尤度					p	R2乗	AICc	BIC
706.78527					6	0.0000	1415.57	1420.52

ロック	追加	パラメータ	推定値	自由度	Wald/ スコアカイ2乗	"p値"
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	切片[Yes]	-0.3697044	1	0	1
<input type="checkbox"/>	<input type="checkbox"/>	客室クラス{18&2-3}	0	1	85.28009	2.6e-20
<input type="checkbox"/>	<input type="checkbox"/>	客室クラス{1-2}	0	1	21.77888	1.87e-5
<input type="checkbox"/>	<input type="checkbox"/>	性別{female-male}	0	1	302.1531	1.1e-67
<input type="checkbox"/>	<input type="checkbox"/>	年齢	0	1	3.042464	0.08111
<input type="checkbox"/>	<input type="checkbox"/>	料金	0	1	64.87624	8e-16

7

変数選択方法を設定

8

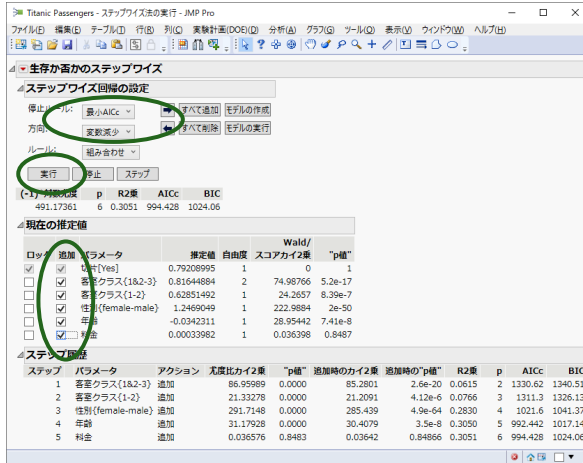
- ▶ 停止ルール
- ▶ 方向
 - ▶ 変数増加法、変数減少法、
変数増減（ステップワイズ）
- ▶ ルール
 - ▶ 3水準以上もつ因子
 - ▶ 効果全体：因子を加えるか否か
 - ▶ 組み合わせ：ネストされた増減
 - ▶ ルールなし：適当な水準で分け、別の因子とみなす

8

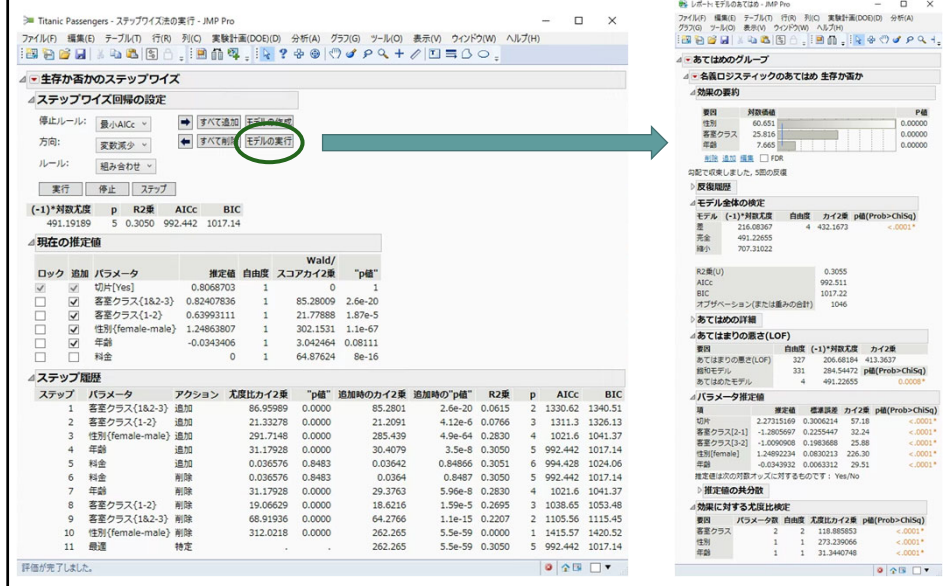
例：変数減少法①

▶ 変数減少法で最小AICのモデルを探索

自分で
「追加」を
クリック
↓
実行



例：変数減少法②



モデルを複数作成

- ▶ 料金を追加
- ▶ モデルの実行

Titantic Passengers - ステップワイズ法の実行 - JMP Pro

ステップワイズ法の実行の設定

停止ルール: 最小AICC

方向: 変数減少

ルール: 組み合わせ

実行 停止 ステップ

(-1)*対数尤度 p R2乗 AICC BIC

491.17361 6 0.3051 994.428 1024.06

現在の推定値

ロック	追加	パラメータ	推定値	自由度	スコアカイ2乗	"p値"
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	切符[Yes]	0.79208995	1	0	1
<input type="checkbox"/>	<input checked="" type="checkbox"/>	客室クラス(1&2-3)	0.81644884	2	74.98766	5.2e-17
<input type="checkbox"/>	<input checked="" type="checkbox"/>	客室クラス(1-2)	0.62851492	1	24.2657	8.39e-7
<input type="checkbox"/>	<input checked="" type="checkbox"/>	性別(female-male)	1.2469049	1	222.9884	2e-50
<input type="checkbox"/>	<input checked="" type="checkbox"/>	年齢	-0.0342311	1	26.95442	7.41e-8
<input type="checkbox"/>	<input checked="" type="checkbox"/>	料金	0.00033962	1	0.036398	0.8487

ステップ履歴

ステップ	パラメータ	アクション	尤度比カイ2乗	"p値"	追加時のカイ2乗	追加時の"p値"	R2乗	p	AICC	BIC
1	客室クラス(1&2-3)	追加	86.95989	0.0000	85.2801	2.6e-20	0.0615	2	1330.62	1340.51
2	客室クラス(1-2)	追加	21.33278	0.0000	21.2091	4.12e-6	0.0766	3	1311.3	1326.13
3	性別(female-male)	追加	291.7148	0.0000	285.439	4.9e-64	0.2830	4	1021.6	1041.37
4	年齢	追加	31.17928	0.0000	30.4079	3.5e-8	0.3050	5	992.442	1017.14
5	料金	追加	0.036576	0.8483	0.03642	0.84866	0.3051	6	994.428	1024.06
6	料金	削除	0.036576	0.8483	0.0364	0.8487	0.3050	5	992.442	1017.14
7	年齢	削除	31.17928	0.0000	29.3763	5.96e-8	0.2830	4	1021.6	1041.37
8	客室クラス(1-2)	削除	19.06629	0.0000	18.6216	1.59e-5	0.2695	3	1038.65	1053.48
9	客室クラス(1&2-3)	削除	68.91936	0.0000	64.2766	1.1e-15	0.2207	2	1105.56	1115.45
10	性別(female-male)	削除	312.0218	0.0000	262.265	5.5e-59	0.0000	1	1415.57	1420.52
11	年齢	削除	31.17928	0.0000	29.3763	5.96e-8	0.2830	4	1021.6	1041.37
12	料金	追加	0.036576	0.8483	64.8762	8e-16	0.3051	6	994.428	1024.06

各モデルのパラメータ推定値

レポート: モデルの推定値 - JMP Pro

あてはめのグループ

名義ロジスティックのあてはめ: 生存か否か

効果の要約

反復確認

モデル全体の検定

あてはめの詳細

あてはまりの長さ (LOF)

パラメータ推定値

項	推定値	標準誤差	カイ2乗	p値(Prob>ChiSq)
切符	2.27315169	0.3006214	57.18	<.0001*
客室クラス(1-1)	-1.2805697	0.2259447	32.24	<.0001*
客室クラス(1-2)	-1.0090008	0.1862688	25.88	<.0001*
性別(female)	1.24892234	0.0830213	228.30	<.0001*
年齢	-0.0343932	0.0063212	29.51	<.0001*

推定値の共分散

効果に対する尤度比検定

名義ロジスティックのあてはめ: 生存か否か

効果の要約

反復確認

モデル全体の検定

あてはめの詳細

あてはまりの長さ (LOF)

パラメータ推定値

項	推定値	標準誤差	カイ2乗	p値(Prob>ChiSq)
切符	2.23705371	0.3489298	41.10	<.0001*
客室クラス(1-1)	-1.2570298	0.2251815	24.27	<.0001*
客室クラス(1-2)	-1.0043828	0.1892059	25.39	<.0001*
性別(female)	1.2469049	0.0835011	222.99	<.0001*
年齢	-0.0342311	0.0063616	28.95	<.0001*
料金	0.00033962	0.017812	0.04	0.8487

推定値の共分散

効果に対する尤度比検定

ROC曲線

▶ ▼ からROC曲線

あてはめのグループ

名義ロジスティックのあてはめ 生存か否か

- 尤度比検定
- Wald検定
- 信頼区間
- オッズ比
- 逆推定...
- 確率の計算式の保存
- 確率の計算式を発行
- ROC曲線**
- リフトチャート
- 混同行列
- 利益行列の指定
- プロファイル

83688 25.88 <.0001*

30213 226.30 <.0001*

どちらの水準を「陽性」としますか?

Yes

No

OK キャンセル

名義ロジスティックのあてはめ 生存か否か

受診者動作特性(ROC)

生存か否か='Yes'を陽性としています。

AUC
0.84006

ROCテーブル

13

閾値を探索

- ▶ ROCテーブルをみる
- ▶ ロジスティック回帰モデルから予測される確率に閾値を設定して感度・特異度を計算

▶ ちょっとわかりにくい

ROCテーブル

確率	1-特異度	感度	感度-(1-特異度)	真陽性	真陰性	偽陽性	偽陰性
.	0.0000	0.0000	0.0000	0	619	0	427
0.9693	0.0016	0.0000	-0.0016	0	618	1	427
0.9544	0.0016	0.0023	0.0007	1	618	1	426
0.9529	0.0016	0.0047	0.0031	2	618	1	425
0.9513	0.0016	0.0117	0.0101	5	618	1	422
0.9497	0.0016	0.0164	0.0148	7	618	1	420
0.9480	0.0016	0.0281	0.0265	12	618	1	415
0.9463	0.0016	0.0351	0.0335	15	618	1	412
0.9427	0.0016	0.0422	0.0405	18	618	1	409
0.9408	0.0016	0.0562	0.0546	24	618	1	403
0.9388	0.0016	0.0656	0.0640	28	618	1	399
0.9368	0.0016	0.0773	0.0757	33	618	1	394
0.9348	0.0032	0.0796	0.0764	34	617	2	393
0.9326	0.0032	0.0843	0.0811	36	617	2	391
0.9304	0.0032	0.0890	0.0858	38	617	2	389

14

料金だけでROC解析

15

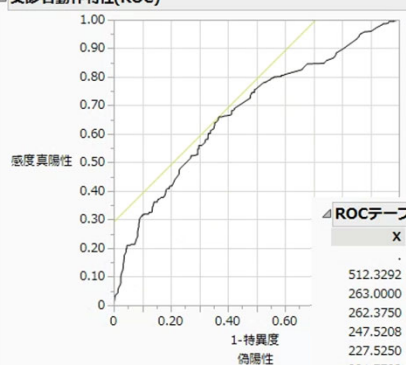
- ▶ 分析> モデルのあてはめ
- ▶ 生存か否かをYに
- ▶ 料金をモデル効果の構成に
 - ▶ 手法は名義ロジスティックのまま
- ▶ ▼からROC曲線

15

確率だけでなく、料金も表示

16

受診者動作特性(ROC)



生存か否か='Yes'を陽性としています。

AUC
0.67462

ROCテーブル

X	確率	1-特異度	感度	感度-(1-特異度)	真陽性	真陰性	偽陽性	偽陰性
.	.	0.0000	0.0000	0.0000	0	808	0	500
512.3292	0.9959	0.0000	0.0080	0.0080	4	808	0	496
263.0000	0.9163	0.0025	0.0160	0.0135	8	806	2	492
262.3750	0.9157	0.0037	0.0280	0.0243	14	805	3	486
247.5208	0.9002	0.0050	0.0320	0.0270	16	804	4	484
227.5250	0.8755	0.0074	0.0380	0.0306	19	802	6	481
221.7792	0.8675	0.0111	0.0400	0.0289	20	799	9	480
211.5000	0.8521	0.0149	0.0440	0.0291	22	796	12	478
211.3375	0.8519	0.0149	0.0520	0.0371	26	796	12	474
164.8667	0.7633	0.0161	0.0580	0.0419	29	795	13	471
153.4625	0.7366	0.0173	0.0620	0.0447	31	794	14	469
151.5500	0.7320	0.0210	0.0680	0.0470	34	791	17	466
146.5208	0.7195	0.0223	0.0720	0.0497	36	790	18	464
136.7792	0.6944	0.0235	0.0740	0.0505	37	789	19	463
135.6333	0.6914	0.0248	0.0800	0.0552	40	788	20	460

16

あてはめの詳細

17

- ▶ RASE : Brierスコアの平方根
 - ▶ Root Average Square Error

指標	学習	定義
エントロピー-R2乗	0.0494	$1 - \text{Loglike}(\text{model}) / \text{Loglike}(0)$
一般化R2乗	0.0865	$(1 - L(0) / L(\text{model}))^{2/n} / (1 - L(0) / L(0))^{2/n}$
平均 -Log p	0.6323	$\sum -\text{Log}(p[j]) / n$
RASE	0.4689	$\sqrt{\sum (y[j] - p[j])^2 / n}$
平均 絶対偏差	0.4407	$\sum y[j] - p[j] / n$
誤分類率	0.3379	$\sum (p[j] * pMax) / n$
N	1308	n

17

練習①

18

- ▶ 料金だけのモデルにおける
相対Brierスコア減少を求めよ
 - ▶ 無情報モデルは
説明変数に何も加えない（切片だけ）モデル

$$\frac{\text{Brier}_{\text{無情報モデル}} - \text{Brier}_{\text{予測モデル}}}{\text{Brier}_{\text{無情報モデル}}}$$

- ▶ 答えは0.0687

18

ROC曲線の重ね合わせ

19

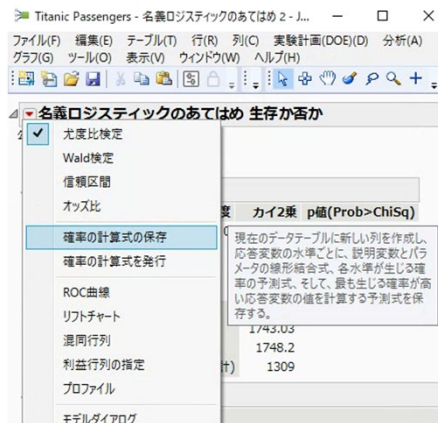
- ▶ 3モデルでのROCを重ね合わせ、AUCを比較しよう
 - ▶ 変数減少法で選ばれたモデル
 - ▶ すべての候補変数を含んだモデル
 - ▶ 料金だけのモデル

19

確率の計算式の保存

20

- ▶ データセットに個人予測確率が生成
 - ▶ 3つのあてはめ結果それぞれで実施



20

データセットに予測確率が追加

変数減少法

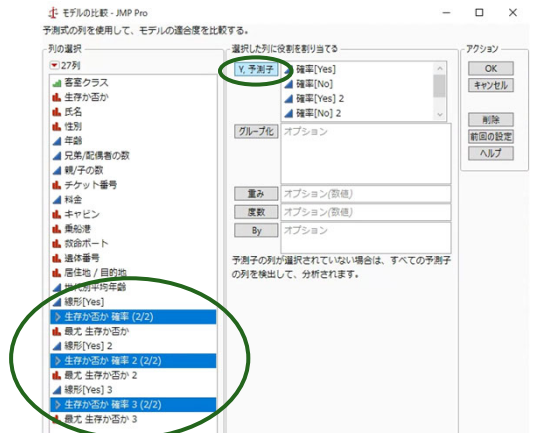
すべて

料金のみ

世代別平均年齢	観形[Yes]	確率[Yes]	確率[No]	観尤 生存か否か	観形[Yes] 2	確率[Yes] 2	確率[No] 2	観尤 生存か否か 2	観形[Yes] 3	確率[Yes] 3	確率[No] 3	観尤 生存か否か 3
27.5	2.5246703305	0.9258533065	0.0741466935	Yes	2.5630731942	0.9284468891	0.0715531109	Yes	1.7493010251	0.8518646191	0.1481353809	Yes
2.5	0.9927010813	0.7296211062	0.2703788938	Yes	1.0102684984	0.730726915	0.2669273085	Yes	1.0047837986	0.7319980903	0.2680019097	Yes
2.5	3.4532875713	0.9693290319	0.0306709681	Yes	3.466995752	0.9697339682	0.0302660318	Yes	1.0047837986	0.7319980903	0.2680019097	Yes
32.5	-0.007567578	0.4981081145	0.5018918855	No	0.0147154817	0.503678804	0.496321196	Yes	1.0047837986	0.7319980903	0.2680019097	Yes
27.5	2.6622432551	0.934761599	0.065238401	Yes	2.679680716	0.9358169489	0.0641830511	Yes	1.0047837986	0.7319980903	0.2680019097	Yes
47.5	-0.626645738	0.3482714949	0.6517285051	No	-0.643921286	0.3443606661	0.6556393339	No	-0.551806681	0.3654453468	0.6345546532	No
62.5	1.3553004717	0.7949948395	0.2050051605	Yes	1.3538916127	0.794765131	0.205234869	Yes	0.0883666817	0.5220773061	0.4779226939	Yes
37.5	-0.317106658	0.4213810388	0.5786189612	No	-0.344863641	0.4146285265	0.5853714735	No	-0.882426499	0.2926752011	0.7073247989	No
52.5	1.6992327831	0.8454345054	0.1545654946	Yes	1.6872044392	0.8438561626	0.1561438374	Yes	-0.241370238	0.4399487048	0.5600512952	No
72.5	-1.417690055	0.1950239656	0.8049760344	No	-1.423436085	0.1941234789	0.8058765211	No	-0.265964368	0.4339881039	0.5661018961	No
47.5	-0.592252507	0.3561181916	0.6438818084	No	-0.541395397	0.3678630371	0.6321369629	No	1.9508794923	0.8755425098	0.1244574902	Yes
17.5	2.9029958731	0.9479943351	0.0520056649	Yes	2.9451159619	0.9500321469	0.0499678531	Yes	1.9508794923	0.8755425098	0.1244574902	Yes
22.5	2.6966364862	0.9368278787	0.0631721213	Yes	2.6859618241	0.9361931837	0.0638068163	Yes	-0.019452737	0.4951369691	0.5048630309	No
27.5	2.6278500239	0.9326325941	0.0673674059	Yes	2.6207449031	0.9321848116	0.0678151884	Yes	0.0994707755	0.5248472098	0.4751527902	Yes
82.5	-1.727229135	0.150942347	0.849057653	No	-1.738143748	0.1495488665	0.8504511335	No	-0.508844784	0.3754643741	0.6243535259	No
.	-0.559589634	0.3636424159	0.6363575841	No

モデルの比較

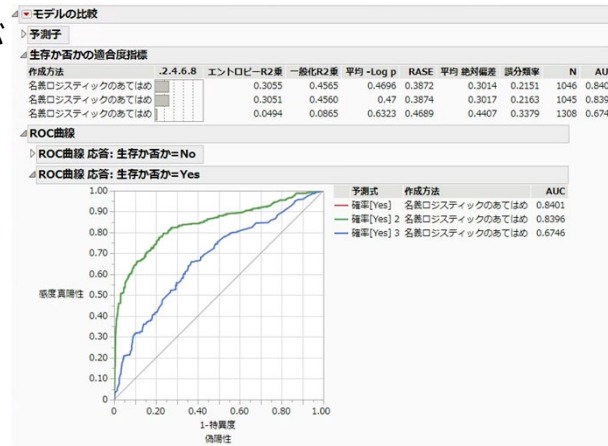
- ▶ 分析 > 予測モデル > モデルの比較
- ▶ ▷ 生存か否か 確率をY, 予測子に
- ▶ 3モデルとも



ROC曲線

- ▶ 変数減少法モデルとすべての候補を含んだモデルがほぼ同じ

- ▶ ROC曲線が重なる



23

AUCの比較

- ▶ AUCの信頼区間やAUC間の検定が実施
- ▶ 変数減少法モデルとすべての候補を含んだモデルはほぼ同じなので検定がうまくできない

予測子	AUC	標準誤差	下側95%	上側95%
確率[Yes]	0.8401	0.0133	0.8123	0.8644
確率[Yes] 2	0.8396	0.0133	0.8117	0.8639
確率[Yes] 3	0.6746	0.0154	0.6437	0.7041

予測子	vs. 予測子	AUCの差	標準誤差	下側95%	上側95%	カイ2乗	p値(Prob>ChiSq)
確率[Yes]	確率[Yes] 2	0.0005					
確率[Yes]	確率[Yes] 3	0.1654	0.0143	0.1374	0.1935	133.54	<.0001*
確率[Yes] 2	確率[Yes] 3	0.1649	0.0143	0.1370	0.1929	133.46	<.0001*

検定	カイ2乗	自由度	p値(Prob>ChiSq)
すべてのAUCが等しい	133.535	1	<.0001*

24

練習② ROC解析

25

- ▶ 以下の2モデルのROC曲線を比較し、AUCの差の検定を実行せよ
 - ▶ 客室クラスだけを用いたモデル
 - ▶ AUC:0.6663
 - ▶ 料金だけを用いたモデル
 - ▶ AUC:0.6746
 - ▶ AUCの差のP値:0.44