

# Department of Statistics

## COURSE STATS 330

### Assignment 4, 2005

Instructions: Hand in your completed assignment to the Student Resource Centre by 4pm on Thursday 29 September.

The file **ICU.txt** contains data measured on patients admitted to an ICU (Intensive Care Unit). The data were collected as part of a study to identify risk factors for mortality in the ICU. The dependent variable is Vital Status, which is a binary variable recording if the patient survived to be discharged from the ICU. The other variables in the data set are described in the list below. For more information on the clinical background to this study, see the article by Lemeshow, Terres, Avrunin and Pastides in the 1988 volume of the Journal of the American Statistical Association, p 348.

#### LIST OF VARIABLES:

Name	Codes/Values	Variable name
Identification Code	ID Number	ID
Vital Status	0 = Lived, 1=died	STA
Age	Years	AGE
Sex	0 = Male, 1=female	SEX
Race	1 = White, 2=Black, 3=other	RACE
Service at ICU Admission	0 = Medical, 1=Surgical	SER
Cancer Part of Present Problem	0 = No, 1=Yes	CAN
History of Chronic Renal Failure	0= No, 1=Yes	CRN
Infection Probable at ICU Admission	0 = No, 1=Yes	INF
CPR Prior to ICU Admission	0 = No, 1=Yes	CPR
Systolic Blood Pressure at Admission	mm Hg	SYS
Heart Rate at ICU Admission	Beats/min	HRA
Previous time in ICU within 6 months	0 = No, 1=Yes	PRE
Type of Admission	0 = Elective, 1 = Emergency	TYP
Serious Fracture	0 = No, 1=Yes	FRA
PO2 from Initial Blood Gases	0 = > 60, 1 = < 60	PO2
PH from Initial Blood Gases	0 => 7.25, 1 =< 7.25	PH
PCO2 from initial Blood Gases	0 = < 45, 1 = > 45	PCO
Bicarbonate from Initial Blood Gases	0 = > 18, 1 = < 18	BIC
Creatinine from Initial Blood Gases	0 = < 2.0, 1 = > 2.0	CRE
Level of Consciousness at ICU Admission	0 = No Coma or Stupor 1 = Deep Stupor, 2=Coma	LOC

1. Read the data into R and make a data frame ICU.df.
2. Using all the variables, fit a logistic regression model to the data which explains the probability of death in terms of the other variables. Using subset selection, or otherwise, identify those variables that seem to have an influence on ICU mortality. Refit the model using these variables, and carefully discuss how they influence mortality. Do any of the explanatory variables in the reduced model require transformation?
3. Subject your model to the usual diagnostic checks. Are there any points that have an influence on the fitted model? What points are they?

NOTE: For the reasons mentioned in class, it is not necessarily a good idea to delete influential points in logistic regression. In this case, retain them.

4. The purpose of the study was to develop a “prognostic index” that would allow the clinicians to predict if an admitted patient will survive or not. A patient is predicted to die if the estimated probability of death exceeds 0.5. How good is your model as a predictor? A claim is made that the predictor (when used as a classifier on new cases) has the following properties

Sensitivity: 42%  
Specificity: 95%  
Percentage correctly classified: 84%

Do you agree with this claim? Hint: use 10-fold cross-validation. I have added a piece of sample code to do cross-validation overleaf. You may need to modify it slightly.

Cross-validation program and data are overleaf (Data also in file ICU.txt)

### **Definitions**

Sensitivity: The probability a patient who dies will be correctly classified

Specificity: The probability a patient who lives will be correctly classified

### Sample program to calculate sensitivity using 10-fold cross-validation

```
n<-length(residuals(ICU.glm)) # n is number of data

nfold<- 10

# divide data into 10 subsamples, each of size m
m<-n%/%nfold

# make vector to store specificity for each sub-sample
Sense<-numeric(nfold)

# put in a formula to describe your model e.g

formula<- STA ~ LOC + TYP + SYS + CAN + AGE

rand.order<-sample(n)
ICU.df<-ICU.df[rand.order,]

# this randomly rearranges the rows of the data frame
sample<-1:m # first pick the first m rows

# now for a loop using each of the subsamples in turn

for(j in 1:nfold){
# fit model using data not in subsample
fit<-glm(formula, family = binomial, data = ICU.df[-sample,])
# use model to predict sample y-values
y.pred<-(predict(fit, ICU.df[sample,],type="response")>0.5)*1
#get actual sample y-values
y.act<-ICU.df$STA[sample]
#calculate proportion of correctly predicted deaths
Sense[j]<-sum(y.act*y.pred)/sum(y.act)
# get new sub-sample
sample<-sample + m
}
# average the nfold results
Av.Sense<- mean(Sense, na.rm=T)
# set na.rm=T in case some samples have no deaths
# if this happens Sense[j] will be 0/0 , so we want to ignore
# these results
```

To get a more accurate answer, you can repeat this several times and average the results.

**Data in file ICU.txt**

ID	STA	AGE	SEX	RACE	SER	CAN	CRN	INF	CPR	SYS	HRA	PRE	TYP	FRA	PO2	PH	PCO	BIC	CRE	LOC
8	0	27	1	1	0	0	0	1	0	142	88	0	1	0	0	0	0	0	0	0
12	0	59	0	1	0	0	0	0	0	112	80	1	1	0	0	0	0	0	0	0
14	0	77	0	1	1	0	0	0	0	100	70	0	0	0	0	0	0	0	0	0
28	0	54	0	1	0	0	0	1	0	142	103	0	1	1	0	0	0	0	0	0
32	0	87	1	1	1	0	0	1	0	110	154	1	1	0	0	0	0	0	0	0
38	0	69	0	1	0	0	0	1	0	110	132	0	1	0	1	0	0	1	0	0
40	0	63	0	1	1	0	0	0	0	104	66	0	0	0	0	0	0	0	0	0
41	0	30	1	1	0	0	0	0	0	144	110	0	1	0	0	0	0	0	0	0
42	0	35	0	2	0	0	0	0	0	108	60	0	1	0	0	0	0	0	0	0
50	0	70	1	1	1	1	0	0	0	138	103	0	0	0	0	0	0	0	0	0
51	0	55	1	1	1	0	0	1	0	188	86	1	0	0	0	0	0	0	0	0
53	0	48	0	2	1	1	0	0	0	162	100	0	0	0	0	0	0	0	0	0
58	0	66	1	1	1	0	0	0	0	160	80	1	0	0	0	0	0	0	0	0
61	0	61	1	1	0	0	1	0	0	174	99	0	1	0	0	1	0	1	1	0
73	0	66	0	1	0	0	0	0	0	206	90	0	1	0	0	0	0	0	1	0
75	0	52	0	1	1	0	0	1	0	150	71	1	0	0	0	0	0	0	0	0
82	0	55	0	1	1	0	0	1	0	140	116	0	0	0	0	0	0	0	0	0
84	0	59	0	1	0	0	0	1	0	48	39	0	1	0	1	0	1	1	0	2
92	0	63	0	1	0	0	0	0	0	132	128	1	1	0	0	0	0	0	0	0
96	0	72	0	1	1	0	0	0	0	120	80	1	0	0	0	0	0	0	0	0
98	0	60	0	1	0	0	0	1	1	114	110	0	1	0	0	0	0	0	0	0
100	0	78	0	1	1	0	0	0	0	180	75	0	0	0	0	0	0	0	0	0
102	0	16	1	1	0	0	0	0	0	104	111	0	1	0	0	0	0	0	0	0
111	0	62	0	1	1	0	1	0	0	200	120	0	0	0	0	0	0	0	0	0
112	0	61	0	1	0	0	0	1	0	110	120	0	1	0	0	0	0	0	0	0
136	0	35	0	1	0	0	0	0	0	150	98	0	1	0	0	0	0	0	0	0
137	0	74	1	1	1	0	0	0	0	170	92	0	0	0	0	0	1	0	0	0
143	0	68	0	1	1	0	0	0	0	158	96	0	0	0	0	0	0	0	0	0
153	0	69	1	1	1	0	0	0	0	132	60	0	1	0	0	0	0	0	0	0
170	0	51	0	1	0	0	0	0	0	110	99	0	1	0	0	0	0	0	0	0
173	0	55	0	1	1	0	0	0	0	128	92	0	0	0	0	0	0	0	0	0
180	0	64	1	3	1	0	0	1	0	158	90	1	1	0	0	0	0	0	0	0
184	0	88	1	1	1	0	0	1	0	140	88	1	1	0	0	0	0	0	0	0
186	0	23	1	1	1	0	0	0	0	112	64	0	1	1	0	0	0	0	0	0
187	0	73	1	1	1	1	0	0	0	134	60	0	0	0	0	0	1	0	0	0
190	0	53	0	3	1	0	0	0	0	110	70	1	0	0	0	0	0	0	0	0
191	0	74	0	1	1	0	0	0	0	174	86	0	0	0	0	0	0	0	0	0
207	0	68	0	1	1	0	0	0	0	142	89	0	0	0	0	0	0	0	0	0
211	0	66	1	1	0	0	0	1	0	170	95	1	1	0	0	0	0	0	0	0
214	0	60	0	1	1	1	0	1	0	110	92	0	0	0	0	0	0	0	0	0
219	0	64	0	1	1	0	0	1	0	160	120	0	0	0	0	0	0	0	0	0
225	0	66	0	2	1	1	0	1	0	150	120	0	0	0	0	0	1	0	0	0
237	0	19	1	1	1	0	0	1	0	142	106	0	1	1	0	0	0	0	0	0
247	0	18	1	1	0	0	0	0	0	146	112	0	1	0	0	0	0	0	0	0
249	0	63	0	1	1	0	0	1	0	162	84	1	1	0	0	0	0	0	0	0
260	0	45	0	1	0	0	0	0	0	126	110	0	1	0	0	0	0	0	0	0
266	0	64	0	1	0	0	0	0	0	162	114	0	1	0	0	0	0	0	0	0
271	0	68	1	1	0	0	0	1	0	200	170	1	1	0	0	0	0	0	0	0
276	0	64	1	1	0	0	0	1	0	126	122	0	1	0	1	0	1	0	0	0
277	0	82	0	1	1	0	0	0	0	135	70	0	0	0	0	0	0	0	0	0
278	0	73	0	1	1	0	0	0	0	170	88	0	0	0	0	0	0	0	0	0
282	0	70	0	1	0	0	0	0	0	86	153	1	1	0	0	0	1	0	0	0
292	0	61	0	1	1	0	0	1	0	68	124	0	1	0	0	0	0	0	0	0
295	0	64	0	1	1	1	0	1	0	116	88	0	0	0	0	0	0	0	0	0
297	0	47	0	1	1	1	0	1	0	120	83	0	0	0	0	0	0	0	0	0
298	0	69	0	1	1	0	0	0	0	170	100	0	0	0	0	0	0	0	0	0
308	0	67	1	1	0	0	0	1	0	190	125	0	1	0	0	0	0	0	0	0
310	0	18	0	1	1	1	0	0	0	156	99	0	0	0	0	0	0	0	0	0

319	0	77	0	1	1	0	0	1	0	158	107	0	0	0	0	0	0	0	0
327	0	32	0	2	1	0	0	0	0	120	84	0	1	0	0	0	0	0	0
333	0	19	1	1	1	0	0	1	0	104	121	1	0	0	0	0	0	0	0
335	0	72	1	1	1	0	0	0	0	130	86	0	1	0	0	0	0	0	0
343	0	49	0	1	0	0	0	1	0	112	112	0	1	0	0	0	0	0	0
357	0	68	1	1	1	0	0	0	0	154	74	0	0	0	0	0	0	0	0
362	0	82	0	1	1	0	1	1	0	130	131	0	1	0	0	0	0	0	0
365	0	32	1	3	0	0	0	1	1	110	118	0	1	0	0	0	0	0	0
369	0	78	1	1	1	0	0	1	0	126	96	0	1	0	0	0	0	0	0
370	0	57	0	1	0	0	0	1	0	128	104	0	1	0	0	0	1	0	0
371	0	46	1	1	1	0	0	0	0	132	90	0	1	0	0	0	0	0	0
376	0	23	0	1	0	0	0	1	0	144	88	0	1	0	0	0	0	0	0
378	0	55	0	1	0	0	0	0	0	132	112	0	1	0	0	0	0	0	0
379	0	18	0	1	1	0	0	0	0	112	76	0	1	1	0	0	0	0	0
381	0	20	0	1	1	0	0	0	0	164	108	0	1	0	0	0	0	0	0
382	0	75	1	1	1	0	0	0	0	100	48	0	0	0	0	0	0	0	0
398	0	79	0	1	1	0	0	1	0	112	67	0	0	0	0	0	0	0	0
401	0	40	0	1	1	0	0	0	0	140	65	0	1	1	0	0	0	0	0
409	0	76	0	1	1	0	0	1	0	110	70	0	1	0	0	0	0	0	0
413	0	66	1	1	1	0	0	1	0	139	92	0	0	0	0	0	0	0	0
416	0	76	0	1	0	0	0	1	0	190	100	0	1	0	0	0	0	0	0
438	0	80	1	1	1	0	0	0	0	162	44	0	1	0	0	0	0	0	0
439	0	23	1	1	0	0	0	1	0	120	88	0	1	0	0	0	0	0	0
440	0	48	0	2	1	0	0	1	0	92	162	1	1	0	0	0	0	0	0
455	0	67	0	2	1	0	0	0	0	90	92	1	0	0	0	0	0	0	0
462	0	69	1	1	1	0	0	0	0	150	85	0	1	0	0	0	0	0	0
495	0	65	0	3	1	0	0	0	0	208	124	0	0	0	0	0	0	0	0
498	0	72	0	1	1	0	0	0	0	126	88	0	0	0	0	0	0	0	0
502	0	55	0	1	0	0	0	0	0	190	136	0	1	0	1	1	1	0	0
505	0	40	0	1	0	0	0	0	0	130	65	0	1	0	0	0	0	0	0
508	0	55	1	1	0	0	0	1	0	110	86	0	1	0	0	0	0	0	0
517	0	34	0	1	1	0	0	0	0	110	80	0	1	1	0	0	0	0	0
522	0	47	1	1	1	0	0	0	0	132	68	0	1	0	0	0	0	0	0
525	0	41	1	1	0	0	0	1	0	118	145	0	1	0	0	1	0	1	0
526	0	84	1	1	0	0	1	1	0	100	103	0	1	0	0	0	0	1	1
546	0	88	1	1	1	0	0	0	0	110	46	1	0	0	0	0	0	0	0
548	0	77	1	1	1	1	0	0	0	212	87	0	0	0	0	0	1	0	0
550	0	80	0	1	0	0	0	0	0	122	126	0	1	0	1	0	0	1	0
552	0	16	0	1	1	0	0	0	0	100	140	0	1	1	0	0	0	0	0
560	0	70	0	1	1	0	0	0	0	160	60	0	0	0	0	0	0	0	0
563	0	83	1	1	1	0	0	1	0	138	91	0	1	0	0	0	0	0	0
573	0	23	0	2	0	0	0	0	0	130	52	0	1	0	0	0	0	0	0
575	0	67	1	1	0	0	0	0	1	120	120	0	1	0	0	1	1	0	0
584	0	18	0	1	1	1	0	0	0	130	140	0	0	0	0	0	0	0	0
597	0	77	1	1	0	0	0	1	0	136	138	0	0	0	1	1	1	0	0
598	0	48	1	1	0	0	0	0	1	128	96	0	1	0	0	0	0	0	0
601	0	24	1	2	0	0	0	0	0	140	86	0	1	0	0	0	0	0	0
605	0	71	1	1	0	0	0	1	0	124	106	0	1	0	0	0	0	0	0
607	0	72	0	1	1	0	0	0	0	134	60	0	1	0	0	0	0	0	0
619	0	77	1	1	1	0	1	0	0	170	115	1	0	0	0	0	0	0	0
620	0	60	0	1	1	0	0	1	0	124	135	0	1	0	0	0	0	0	0
639	0	46	0	1	1	1	0	0	0	110	128	0	0	0	0	0	0	0	0
644	0	65	1	1	0	0	0	0	0	100	105	0	1	0	0	0	0	0	0
645	0	36	0	1	0	0	0	0	0	224	125	0	1	0	0	0	0	0	0
648	0	68	0	1	1	0	0	0	0	112	64	0	0	0	0	0	0	0	0
655	0	58	0	1	0	0	0	0	0	154	98	0	1	0	0	0	0	0	0
659	0	76	1	1	0	0	0	1	0	92	112	0	1	0	0	0	0	0	0
669	0	41	1	2	0	0	0	0	0	110	144	0	1	0	0	0	0	1	1
670	0	20	0	3	0	0	0	0	0	120	68	0	1	0	0	0	0	0	0
674	0	91	0	1	0	0	1	1	0	152	125	0	1	0	0	0	0	0	0
675	0	75	0	1	1	0	0	0	0	140	90	0	1	0	0	0	0	0	0
676	0	25	1	1	0	0	0	0	0	131	135	0	1	0	0	0	0	1	0
709	0	70	0	1	0	0	0	1	0	78	143	0	1	0	1	0	0	0	0

713	0	47	0	1	1	0	0	0	0	156	112	0	1	0	0	0	0	0	0
727	0	75	0	3	1	0	0	0	0	144	120	0	1	0	0	0	0	0	1
728	0	40	0	2	0	0	0	0	1	160	150	1	1	1	0	0	0	0	0
732	0	71	0	1	0	0	0	0	1	148	192	0	1	0	1	1	1	0	0
746	0	70	1	1	0	0	0	1	0	90	140	0	1	0	1	0	0	1	0
749	0	58	0	1	1	0	0	0	0	148	95	1	1	0	0	0	0	0	0
754	0	54	0	1	1	0	0	0	0	136	80	0	0	0	0	0	0	0	0
761	0	77	0	1	1	0	0	0	0	128	59	0	0	0	0	0	0	0	0
763	0	55	0	1	1	1	0	1	0	138	140	0	0	0	0	0	0	0	0
764	0	21	0	1	1	0	0	0	0	120	62	0	1	0	0	0	0	0	0
765	0	53	0	2	0	0	1	0	1	170	115	0	1	0	0	0	0	0	0
766	0	31	1	1	0	1	1	1	1	146	100	0	1	0	0	1	1	0	0
772	0	71	0	1	1	1	0	0	0	204	52	0	0	0	0	0	0	0	0
776	0	49	0	2	0	0	0	0	0	150	100	0	1	0	0	0	0	0	0
784	0	60	1	2	0	0	0	1	0	116	92	1	1	0	0	0	0	0	0
794	0	50	0	1	0	0	0	1	0	156	99	0	1	0	1	0	1	0	0
796	0	45	1	1	1	0	0	0	0	132	109	0	1	1	0	0	0	0	0
809	0	21	0	1	1	0	0	0	0	110	90	0	1	0	0	0	0	0	0
814	0	73	1	1	1	0	0	0	0	130	83	0	1	0	0	0	0	0	0
816	0	28	0	1	1	0	0	1	0	122	80	1	0	1	0	0	0	0	0
829	0	17	0	1	1	0	0	0	0	140	78	0	1	1	0	0	0	0	0
837	0	17	1	3	0	0	0	0	0	130	140	0	1	0	0	0	0	0	0
846	0	21	1	1	1	0	0	0	0	142	79	0	1	0	0	0	0	0	0
847	0	68	1	1	1	1	0	0	0	91	79	0	0	0	0	0	0	0	0
863	0	17	0	3	1	0	0	0	0	136	78	0	1	0	0	0	0	0	0
867	0	60	0	1	0	0	0	1	0	108	120	0	1	0	0	0	0	0	0
875	0	69	0	1	1	0	0	0	0	169	73	0	1	0	0	0	0	0	0
877	0	88	1	1	0	0	1	0	0	190	88	0	1	0	0	0	0	0	0
880	0	20	0	1	1	0	0	0	0	120	80	0	1	0	0	0	0	0	0
881	0	89	1	1	1	0	0	0	0	190	114	0	1	0	0	0	1	0	2
889	0	62	1	1	0	0	0	0	0	110	78	0	1	0	0	0	0	0	0
893	0	46	0	1	0	0	1	1	0	142	89	0	1	0	0	1	0	1	0
906	0	19	0	1	1	0	0	1	0	100	137	0	1	0	0	0	0	0	0
912	0	71	0	1	0	0	0	1	0	124	124	0	1	0	1	1	1	0	0
915	0	67	0	1	1	0	0	0	0	152	78	0	0	0	0	0	0	0	0
923	0	20	0	1	1	0	0	0	0	104	83	0	1	0	0	0	0	0	0
924	0	73	1	2	0	0	1	0	0	162	100	0	1	0	0	0	0	0	0
925	0	59	0	1	0	0	0	0	0	100	88	0	1	0	0	0	0	0	0
929	0	42	0	1	1	0	0	0	0	122	84	0	1	1	0	0	0	0	0
4	1	87	1	1	1	0	0	1	0	80	96	0	1	1	1	1	1	0	0
27	1	76	1	1	1	0	0	1	0	128	90	1	1	0	0	0	0	0	0
47	1	78	0	1	0	0	0	1	0	130	132	0	1	0	0	0	0	1	0
52	1	63	0	1	0	0	1	1	0	112	106	1	1	0	1	0	0	0	0
127	1	19	0	1	1	0	0	0	0	140	76	0	1	0	0	0	0	0	0
145	1	67	1	1	0	0	0	1	0	62	145	0	1	0	0	0	0	0	1
154	1	53	1	1	0	0	0	1	0	148	128	0	1	0	0	1	1	0	0
165	1	92	0	1	0	0	0	1	0	124	80	0	1	0	0	0	0	1	0
195	1	57	0	1	0	0	0	1	1	110	124	0	1	0	0	0	0	0	2
202	1	75	1	1	1	1	0	0	0	130	136	0	0	0	0	0	0	0	0
204	1	91	0	1	0	0	0	1	0	64	125	0	1	0	0	0	1	0	0
208	1	70	0	1	1	0	0	0	0	168	122	0	0	0	1	0	0	0	1
222	1	88	0	1	0	0	0	1	1	141	140	0	1	0	0	0	0	0	0
238	1	41	0	1	1	0	0	1	0	140	58	0	1	0	0	0	0	0	2
241	1	61	0	1	0	0	0	0	0	140	81	0	1	0	0	0	0	0	0
273	1	80	0	1	1	0	0	0	0	100	85	0	1	0	0	0	0	0	0
285	1	40	0	1	0	0	0	1	0	86	80	1	1	0	0	0	0	0	0
299	1	75	0	1	0	0	0	1	0	90	100	0	1	0	0	0	0	0	1
331	1	63	1	1	1	0	1	1	1	36	86	0	1	1	0	0	0	0	2
346	1	75	1	1	0	1	0	0	0	190	94	0	1	0	0	0	0	0	0
380	1	20	0	1	1	0	0	0	0	148	72	0	1	1	0	0	0	0	0
384	1	71	0	1	0	0	0	0	0	142	95	0	1	0	0	0	0	0	0
412	1	51	1	1	1	0	0	1	0	134	100	1	1	0	0	0	0	0	1
427	1	65	0	1	0	0	0	0	0	66	94	0	1	0	0	0	0	0	2

442	1	69	1	3	0	0	1	0	0	170	60	1	1	0	1	0	0	0	0
461	1	55	0	1	1	0	1	1	0	122	100	1	1	0	0	0	0	0	0
468	1	50	1	1	1	1	0	0	0	120	96	0	1	0	0	0	0	0	0
490	1	78	0	1	0	0	0	1	0	110	81	0	1	0	0	0	0	0	0
518	1	71	1	1	0	0	0	0	1	70	112	0	1	0	0	0	0	0	2
611	1	85	1	1	1	0	0	0	0	136	96	0	1	0	0	0	0	0	0
613	1	75	0	1	0	0	1	1	0	130	119	0	1	0	0	1	0	1	0
666	1	65	1	1	0	0	0	1	1	104	150	0	1	0	0	0	1	0	2
671	1	49	0	1	0	0	0	1	1	140	108	0	1	0	0	0	0	1	0
706	1	75	1	1	0	0	1	1	1	150	66	0	1	0	0	0	0	0	2
740	1	72	1	1	0	0	0	0	0	90	160	0	1	0	0	0	0	0	0
751	1	69	0	1	0	0	1	0	0	80	81	0	1	0	0	0	0	0	2
752	1	64	0	1	0	1	0	1	0	80	118	0	1	0	1	0	0	0	0
789	1	60	0	1	0	0	0	1	0	56	114	1	1	0	0	1	0	1	0
871	1	60	0	3	1	0	1	1	0	130	55	0	1	0	0	0	0	0	1
921	1	50	1	2	0	0	0	0	0	256	64	0	1	0	0	0	0	0	1