

Classification results

Stat 543

5-28-15

①

		Predicted	
		0	1
Actual (observed)	0	True negative	False positive
	1	False negative	True positive

$$ACC = \text{accuracy rate} = \frac{\text{True neg} + \text{True pos}}{\text{Total}}$$

$$FPR = \text{false positive rate} = \text{fall out rate} = \frac{\text{False pos}}{\text{Actual 0}}$$

$$FNR = \text{false negative rate} = \frac{\text{False neg}}{\text{Actual 1}}$$

$$TPR = \text{true positive rate} = \text{sensitivity} = \frac{\text{True pos.}}{\text{Actual 1}} \quad (2)$$

$$TNR = \text{true negative rate} = \text{specificity} = \frac{\text{True neg}}{\text{Actual 0}}$$

$$FNR + TPR = 1$$

$$FPR + TNR = 1$$

From the discrim. analy. handout,

$$ACC = \frac{286 + 51}{362} = 93\%$$

$$TPR = \frac{51}{59} = 86.4\%$$

$$TNR = \frac{286}{303} = 94.4\%$$

③

From the logistic reg handout

$$ACC = \frac{297 + 43}{362} = 93.9\%$$

$$TPR = \frac{43}{59} = 72.9\% \quad (\text{sens.})$$

$$TNR = \frac{297}{303} = 98.0\% \quad (\text{spec.})$$

Sampling

④

Simple random sample

- Every item in the population has the same chance of being selected
- Every pair of items has the same chance as every other pair
- Every triple
- etc.

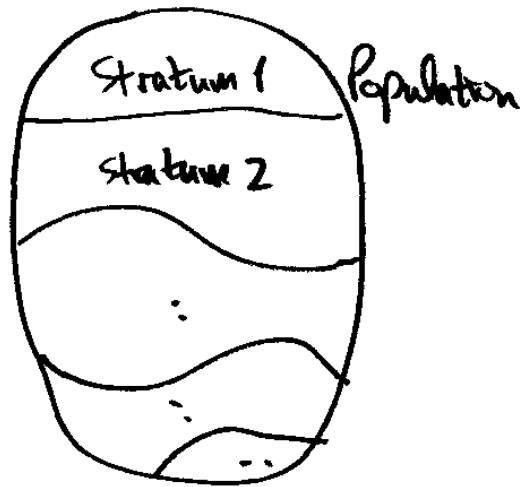
Systematic sampling

- Select a random starting point
 - Select a skip interval
- Ex: 7th person and every 20th person after that.

Stratified sampling

(5)

Select a
simple random
sample from
each stratum

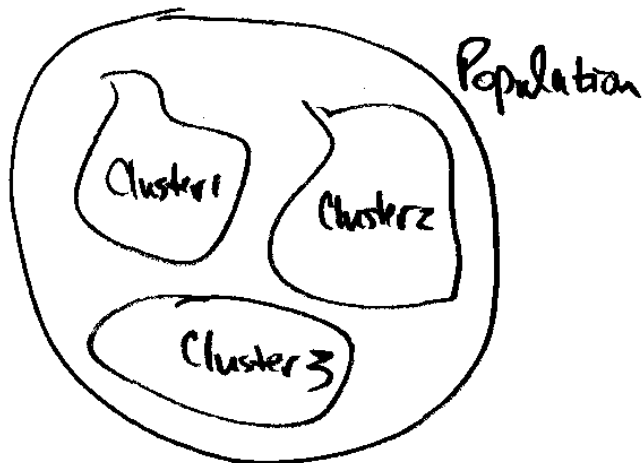


proportional allocation: sample the same % from
each stratum

Cluster sampling

(6)

Assume the
clusters are
very similar
to each other.



If so, then you
can choose just one cluster.

```

WEIGHT BY Freq.
DISCRIMINANT
  /GROUPS=Death(0 1)
  /VARIABLES=Aggrlev VicRace
  /ANALYSIS ALL
  /PRIORS EQUAL
  /STATISTICS=RAW TABLE
  /CLASSIFY=NONMISSING POOLED.

```

Discriminant

Analysis Case Processing Summary

Unweighted Cases N Percent

		N	Percent
Valid		21	100.0
Excluded	Missing or out-of-range group codes	0	.0
	At least one missing discriminating variable	0	.0
	Both missing or out-of-range group codes and at least one missing discriminating variable	0	.0
	Total	0	.0
Total		21	100.0

Group Statistics

D		Valid N (listwise)	
		Unweighted	Weighted
0	A	9	303.000
	V	9	303.000
1	A	12	59.000
	V	12	59.000
Total	A	21	362.000
	V	21	362.000

Analysis 1

Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.759 ^a	100.0	100.0	.798

a.

sis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.362	364.356	2	.000

Standardized Canonical Discriminant Function Coefficients

	Function
	1
A	.957
V	.222

Structure Matrix

	Function
	1
A	.975
V	.301

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.

**Canonical
Discriminant Function
Coefficients**

	Function
	1
A	1.041
V	.497
(Constant)	-2.111

Unstandardized
coefficients

**Functions at
Group Centroids**

D	Function
	1
0	-.584
1	2.997

Unstandardized
canonical
discriminant
functions
evaluated at group
means

Classification Statistics

Classification Processing Summary

Processed	21	
Excluded	Missing or out-of-range group codes	0
	At least one missing discriminating variable	0
Used in Output	21	

Prior Probabilities for Groups

D	Prior	Cases Used in Analysis	
		Unweighted	Weighted
0	.500	9	303.000
1	.500	12	59.000
Total	1.000	21	362.000

Classification Results^a

			Predicted Group Membership		Total
			0	1	
Original	Count	0	286	17	303
		1	8	51	59
	%	0	94.4	5.6	100.0
		1	13.6	86.4	100.0

a.

>Warning # 3211

>On at least one case, the value of the weight variable was zero, negative, or
>missing. Such cases are invisible to statistical procedures and graphs which
>need positively weighted cases, but remain on the file and are processed by
>non-statistical facilities such as LIST and SAVE.

LOGISTIC REGRESSION VARIABLES Death

/METHOD=ENTER Aggrlev VicRace

/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

Logistic Regression

[DataSet1] F:\logistic.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	21	100.0
	Missing Cases	0	.0
	Total	21	100.0
Unselected Cases		0	.0
Total		21	100.0

a.

number of cases.

Dependent Variable Encoding

Original Value	Internal Value
0	0
1	1

Block 0: Beginning Block

Classification Table^{a,b}

Observed			Predicted		
			D		Percentage Correct
			0	1	
Step 0	D	0	303	0	100.0
		1	59	0	.0
Overall Percentage					83.7

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	-1.636	.142	132.208	1	.000	.195

Variables not in the Equation

	Score	df	Sig.
Step 0 Variables Aggriev	226.571	1	.000
VicRace	49.888	1	.000
Overall Statistics	230.800	2	.000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1 Step	208.402	2	.000
Block	208.402	2	.000
Model	208.402	2	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	113.477 ^a	.438	.743

a.

estimates changed by less than .001.

Classification Table^a

Observed			Predicted		
			D		Percentage Correct
			0	1	
Step 1	D	0	297	6	98.0
		1	16	43	72.9
Overall Percentage					93.9

a. The cut value is .500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a Aggrlev	1.540	.187	67.989	1	.000	4.663
VicRace	1.811	.536	11.406	1	.001	6.114
Constant	-6.676	.757	77.683	1	.000	.001

a. Variable(s) entered on step 1: Aggrlev, VicRace.

>Warning # 3211

>On at least one case, the value of the weight variable was zero, negative, or
>missing. Such cases are invisible to statistical procedures and graphs which
>need positively weighted cases, but remain on the file and are processed by
>non-statistical facilities such as LIST and SAVE.