

```

* =====

* save this code as "script logit accuracy by hand.do"

cap drop predictedclass TN FN FP TP sum_TN sum_FN sum_FP sum_TP recall_0 recall_1 precision_0
precision_1 fscore_0 fscore_1 accuracy_0 accuracy_1 accuracy_t

gen predictedclass = (predict_prob>=cutoff) if predict_prob!=.

gen TN = (actual==0 & predictedclass==0)      if predict_prob!=.
gen FN = (actual==1 & predictedclass==0)      if predict_prob!=.
gen FP = (actual==0 & predictedclass==1)      if predict_prob!=.
gen TP = (actual==1 & predictedclass==1)      if predict_prob!=.

egen sum_TP = sum(TP)
egen sum_FN = sum(FN)
egen sum_TN = sum(TN)
egen sum_FP = sum(FP)

gen recall_0    = 100*(sum_TN/(sum_TN + sum_FP))
gen recall_1    = 100*(sum_TP/(sum_TP + sum_FN))
gen precision_0 = 100*(sum_TN/(sum_FN + sum_TN))
gen precision_1 = 100*(sum_TP/(sum_FP + sum_TP))
gen fscore_0    = 2*((recall_0*precision_0)/(recall_0+precision_0))
gen fscore_1    = 2*((recall_1*precision_1)/(recall_1+precision_1))
gen accuracy_0  = 100*((sum_TN)/(sum_TP + sum_TN + sum_FP + sum_FN))
gen accuracy_1  = 100*((sum_TP)/(sum_TP + sum_TN + sum_FP + sum_FN))
gen accuracy_t  = 100*((sum_TP + sum_TN)/(sum_TP + sum_TN + sum_FP + sum_FN))

* =====

```

* =====

* save this code as "prediction logit.do"

clear all

* set working directory

*global mydata "G:/My Drive/Lectures/MSc applied micro/Lecture_7"

*global mydata "/Users/nicholasgiannakopoulos/Library/CloudStorage/GoogleDrive-giannakopoulosnicholas@gmail.com/My Drive/Lectures/MSc applied micro/Lecture_7"

use <http://www.stata.com/data/jwooldridge/eacsap/mroz>, clear

tab inlf

// Fit the logistic regression model

* =====

// Model 1

logit inlf i.age

cap drop logitm1 actual

edit inlf

predict logitm1, pr

estat classification

*roctab inlf logitm1

cap drop predict_prob

gen predict_prob = logitm1

cap drop actual

gen actual = inlf

// Get confusion table

```

* cutoff point

cap drop cutoff

gen cutoff = 0.5

* performance measures

qui do "${mydata}/script logit accuracy by hand.do"

tabstat recall_0 recall_1 precision_0 precision_1 fscore_0 fscore_1 accuracy_0 accuracy_1 accuracy_t in
1, col(stats) format(%4.2f)


preserve

keep recall_0 recall_1 precision_0 precision_1 fscore_0 fscore_1 accuracy_0 accuracy_1 accuracy_t

keep in 1

gen model=1

save "${mydata}/model1.dta", replace

restore


* =====

// Model 2

logit inlf i.age i.educ

cap drop logitm1 actual

predict logitm1, pr

estat classification

*roctab inlf logitm1

cap drop predict_prob

gen predict_prob = logitm1

cap drop actual

gen actual = inlf

// Get confusion table

* cutoff point

cap drop cutoff

```

```

gen cutoff = 0.5

* performance measures

qui do "${mydata}/script logit accuracy by hand.do"

tabstat recall_0 recall_1 precision_0 precision_1 fscore_0 fscore_1 accuracy_0 accuracy_1 accuracy_t in
1, col(stats) format(%4.2f)


preserve

keep recall_0 recall_1 precision_0 precision_1 fscore_0 fscore_1 accuracy_0 accuracy_1 accuracy_t

keep in 1

gen model=2

save "${mydata}/model2.dta", replace

restore


* =====

// Model 3

logit inlf i.age i.educ i.kidslt6

cap drop logitm1 actual

predict logitm1, pr

estat classification

*roctab inlf logitm1

cap drop predict_prob

gen predict_prob = logitm1

cap drop actual

gen actual = inlf

// Get confusion table

* cutoff point

cap drop cutoff

gen cutoff = 0.5

```

* performance measures

```
qui do "${mydata}/script logit accuracy by hand.do"
```

```
tabstat recall_0 recall_1 precision_0 precision_1 fscore_0 fscore_1 accuracy_0 accuracy_1 accuracy_t in  
1, col(stats) format(%4.2f)
```

```
preserve
```

```
keep recall_0 recall_1 precision_0 precision_1 fscore_0 fscore_1 accuracy_0 accuracy_1 accuracy_t
```

```
keep in 1
```

```
gen model=3
```

```
save "${mydata}/model3.dta", replace
```

```
restore
```

```
* =====
```

```
// Model 4
```

```
logit inlf i.age i.educ i.kidslt6 i.kidsge6
```

```
cap drop logitm1 actual
```

```
predict logitm1, pr
```

```
estat classification
```

```
*roctab inlf logitm1
```

```
cap drop predict_prob
```

```
gen predict_prob = logitm1
```

```
cap drop actual
```

```
gen actual = inlf
```

```
// Get confusion table
```

```
* cutoff point
```

```
cap drop cutoff
```

```
gen cutoff = 0.5
```

* performance measures

qui do "\${mydata}/script logit accuracy by hand.do"

tabstat recall_0 recall_1 precision_0 precision_1 fscore_0 fscore_1 accuracy_0 accuracy_1 accuracy_t in
1, col(stats) format(%4.2f)

preserve

keep recall_0 recall_1 precision_0 precision_1 fscore_0 fscore_1 accuracy_0 accuracy_1 accuracy_t

keep in 1

gen model=4

save "\${mydata}/model4.dta", replace

restore

* =====

// Model 5

logit inlf i.age i.educ i.kidslt6 i.kidsge6 c.nwifeinc

cap drop logitm1 actual

predict logitm1, pr

estat classification

*roctab inlf logitm1

cap drop predict_prob

gen predict_prob = logitm1

cap drop actual

gen actual = inlf

// Get confusion table

* cutoff point

cap drop cutoff

gen cutoff = 0.5

* performance measures

```

qui do "${mydata}/script logit accuracy by hand.do"

tabstat recall_0 recall_1 precision_0 precision_1 fscore_0 fscore_1 accuracy_0 accuracy_1 accuracy_t in
1, col(stats) format(%4.2f)

preserve

keep recall_0 recall_1 precision_0 precision_1 fscore_0 fscore_1 accuracy_0 accuracy_1 accuracy_t

keep in 1

gen model=5

save "${mydata}/model5.dta", replace

restore

* =====

// Model 6

logit inlf i.age i.educ i.kidslt6 i.kidsge6 c.nwifeinc i.motheeduc i.fatheduc

cap drop logitm1 actual

predict logitm1, pr

estat classification

*roctab inlf logitm1

cap drop predict_prob

gen predict_prob = logitm1

cap drop actual

gen actual = inlf

// Get confusion table

* cutoff point

cap drop cutoff

gen cutoff = 0.5

* performance measures

qui do "${mydata}/script logit accuracy by hand.do"

```

```
tabstat recall_0 recall_1 precision_0 precision_1 fscore_0 fscore_1 accuracy_0 accuracy_1 accuracy_t in
1, col(stats) format(%4.2f)
```

```
preserve
```

```
keep recall_0 recall_1 precision_0 precision_1 fscore_0 fscore_1 accuracy_0 accuracy_1 accuracy_t
```

```
keep in 1
```

```
gen model=6
```

```
save "${mydata}/model6.dta", replace
```

```
restore
```

```
* =====
```

```
// Model 7
```

```
logit inlf i.age i.educ i.kidslt6 i.kidsge6 c.nwifeinc i.motheeduc i.fatheduc i.city
```

```
cap drop logitm1 actual
```

```
predict logitm1, pr
```

```
estat classification
```

```
*roctab inlf logitm1
```

```
cap drop predict_prob
```

```
gen predict_prob = logitm1
```

```
cap drop actual
```

```
gen actual = inlf
```

```
// Get confusion table
```

```
* cutoff point
```

```
cap drop cutoff
```

```
gen cutoff = 0.5
```

```
* performance measures
```

```
qui do "${mydata}/script logit accuracy by hand.do"
```



```
tabstat recall_0 recall_1 precision_0 precision_1 fscore_0 fscore_1 accuracy_0 accuracy_1 accuracy_t in
1, col(stats) format(%4.2f)
```

```
preserve
```

```
keep recall_0 recall_1 precision_0 precision_1 fscore_0 fscore_1 accuracy_0 accuracy_1 accuracy_t
```

```
keep in 1
```

```
gen model=7
```

```
save "${mydata}/model7.dta", replace
```

```
restore
```

```
*=====
```

```
* compare performance measures
```

```
use "${mydata}/model1.dta", clear
```

```
append using "${mydata}/model2.dta"
```

```
append using "${mydata}/model3.dta"
```

```
append using "${mydata}/model4.dta"
```

```
append using "${mydata}/model5.dta"
```

```
append using "${mydata}/model6.dta"
```

```
append using "${mydata}/model7.dta"
```

```
order model recall_0 recall_1 precision_0 precision_1 fscore_0 fscore_1 accuracy_0 accuracy_1
accuracy_t
```

```
list
```

```
twoway ///
```

```
(connected recall_0 model, mcolor(maroon) lcolor(maroon)) ///
```

```
(connected recall_1 model, mcolor(navy) lcolor(navy)) ///
```

```
, ylab(0(10)100, angle(0) format(%4.0f)) scheme(s1color) ///
```

```
xlab(1 "M1" 2 "M2" 3 "M3" 4 "M4" 5 "M5" 6 "M6" 7 "M7") ///
```

```
yti("%") xti("Model Specification") ti("Recall rate") ///  
legend(order(1 "Y=0" 2 "Y=1"))
```

```
twoway ///  
(connected precision_0 model, mcolor(maroon) lcolor(maroon)) ///  
(connected precision_1 model, mcolor(navy) lcolor(navy)) ///  
, ylab(0(10)100, angle(0) format(%4.0f)) scheme(s1color) ///  
xlab(1 "M1" 2 "M2" 3 "M3" 4 "M4" 5 "M5" 6 "M6" 7 "M7") ///  
yti("%") xti("Model Specification") ti("Precision rate") ///  
legend(order(1 "Y=0" 2 "Y=1"))
```

```
twoway ///  
(connected fscore_0 model, mcolor(maroon) lcolor(maroon)) ///  
(connected fscore_1 model, mcolor(navy) lcolor(navy)) ///  
, ylab(0(10)100, angle(0) format(%4.0f)) scheme(s1color) ///  
xlab(1 "M1" 2 "M2" 3 "M3" 4 "M4" 5 "M5" 6 "M6" 7 "M7") ///  
yti("%") xti("Model Specification") ti("F-Score") ///  
legend(order(1 "Y=0" 2 "Y=1"))
```

```
twoway ///  
(connected accuracy_0 model, mcolor(maroon) lcolor(maroon)) ///  
(connected accuracy_1 model, mcolor(navy) lcolor(navy)) ///  
, ylab(0(10)100, angle(0) format(%4.0f)) scheme(s1color) ///  
xlab(1 "M1" 2 "M2" 3 "M3" 4 "M4" 5 "M5" 6 "M6" 7 "M7") ///  
yti("%") xti("Model Specification") ti("Accuracy rate") ///  
legend(order(1 "Y=0" 2 "Y=1"))
```