

# MEASURING THE FIT OF THE MODEL USING LOG LIKELIHOODS

```
--> READ; NREC=10;NVAR=2; FILE=S206L.ASC; FORMAT=(2F8.2);
      NAMES(X1=UNEMP, X2=GENDER)$
--> OPEN; OUTPUT=S206L.OUT$
--> LIST; UNEMP, GENDER$
```

Listing of raw data (Current sample)

Line	Observ.	UNEMP	GENDER
1	1	.00000	1.0000
2	2	1.0000	.00000
3	3	1.0000	1.0000
4	4	1.0000	1.0000
5	5	1.0000	1.0000
6	6	.00000	.00000
7	7	.00000	.00000
8	8	.00000	.00000
9	9	.00000	1.0000
10	10	.00000	.00000

	Male (0)	Female (1)	
Employed (0)	4	2	6
Unemployed (1)	1	3	4
	5	5	10

$$CHI-SQUARED = (4-3)^2/3 + (2-3)^2/3 + (1-2)^2/2 + (3-2)^2/2 = 1/3 + 1/3 + 1/2 + 1/2 = 2/3 + 1 = 1.6667$$

ODDS OF BEING UNEMPLOYED

MALE

$$1/4 = .2 / (1 - .2)$$

FEMALE

$$3/2 = .6 / (1 - .6)$$

$$(3/2) / (1/4) = 1.5 / .25 = 6 \quad \text{THE RATIO OF THE ODDS (ODDS RATIO)}$$

LOG OF THE ODDS OF BEING UNEMPLOYED

MALE

$$\text{LOG}(1/4) = \text{LOG}(.25) = -1.386294$$

LOG OF THE ODDS OF BEING UNEMPLOYED

FEMALE

$$\text{LOG}(3/2) = \text{LOG}(1.5) = 0.4054651$$

$$\text{THE DIFFERENCE IS } 0.4054651 - (-1.386294) = 1.791759 \quad \text{---} \rightarrow e^{1.791759} = 6$$

```
--> LOGIT; LHS=UNEMP;RHS=ONE,GENDER$
Normal exit from iterations. Exit status=0.
```

```
-----+-----
| Binary Logit Model for Binary Choice
| Maximum Likelihood Estimates
| Model estimated: Feb 05, 2011 at 02:07:51AM.
| Dependent variable          UNEMP
| Weighting variable          None
| Number of observations       10
| Iterations completed        5
| Log likelihood function      -5.867070
| Number of parameters         2
| Info. Criterion: AIC =       1.57341
|   Finite Sample: AIC =       1.74484
| Info. Criterion: BIC =       1.63393
| Info. Criterion:HQIC =       1.50703
| Restricted log likelihood    -6.730117
| McFadden Pseudo R-squared   .1282364
| Chi squared                  1.726092
| Degrees of freedom           1
| Prob[ChiSq > value] =       .1889107
|-----+-----
```

This Chi squared is the same as above except it has a correction for continuity. The table above has counts and those can take only discrete (natural) numbers.

```
-----+-----+-----+-----+-----+-----+
| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
|-----+-----+-----+-----+-----+-----+
| Constant | -1.386294367 | 1.1180340 | -1.240 | .2150 |
| GENDER   | 1.791759469 | 1.4433757 | 1.241 | .2145 | .50000000
|-----+-----+-----+-----+-----+-----+
| Characteristics in numerator of Prob[Y = 1]
|-----+-----+-----+-----+-----+-----+
| CALCULATING THE BASE LINE: RESTRICTED LOG LIKELIHOOD
```



WHAT IS THE PROBABILITY OF BEING UNEMPLOYED GIVEN THAT YOU ARE MALE ?

1 OF 5 MALES ARE UNEMPLOYED  $1/5=.2$  THE CONDITIONAL PROBABILITY

ARE		Male (0)	Female (1)	
YOU	Employed (0)	4	2	6
	Unemployed (1)	1	3	4
ARE		5	5	10

OF BEING UNEMPLOYED GIVEN THAT YOU  
 MALE EQUALS  $1/5=.2$   $Pr(U/M)=.2$   
 OF BEING NOT UNEMPLOYED GIVEN THAT  
 ARE A MALE EQUALS  $4/5=.8$   $Pr(NU/M)=.8$   
 OF BEING UNEMPLOYED GIVEN THAT YOU  
 FEMALE EQUALS  $3/5=.6$   $Pr(U/F)=.6$   
 OF BEING NOT UNEMPLOYED GIVEN THAT

YOU ARE A FEMALE EQUALS  $2/5=.4$   $Pr(NU/F)=.4$

$NU \ U \ U \ U \ U \ NU \ NU \ NU \ NU \ NU$   
 $F \ M \ F \ F \ F \ M \ M \ M \ F \ M$   
 $.4^* \ .2^* \ .6^* \ .6^* \ .6^* \ .8^* \ .8^* \ .8^* \ .4^* \ .8 = .4^2 * .6^3 * .2^1 * .8^4 = 0.0028311552$

$LOG(0.0028311552) = -5.867070$

THE LIKELIHOOD IS NOW 0.0028311552 OR .283% THE LOG LIKELIHOOD IS -5.867070.  
**COMPARE** IT TO THE LIKELIHOOD OF .001194394 OR .119% AND LOG LIKELIHOOD OF -6.730117

--> LOGIT; LHS=UNEMP;RHS=ONE,GENDER\$

Normal exit from iterations. Exit status=0.

```

+-----+
| Binary Logit Model for Binary Choice |
| Maximum Likelihood Estimates         |
| Model estimated: Feb 05, 2011 at 02:07:51AM. |
| Dependent variable                   UNEMP |
| Weighting variable                   None  |
| Number of observations                10   |
| Iterations completed                 5    |
| Log likelihood function               -5.867070 |
| Number of parameters                 2    |
| Info. Criterion: AIC =                1.57341 |
|   Finite Sample: AIC =                1.74484 |
| Info. Criterion: BIC =                1.63393 |
| Info. Criterion:HQIC =               1.50703 |
| Restricted log likelihood             -6.730117 |
| McFadden Pseudo R-squared           .1282364 |
| Chi squared                          1.726092 |
| Degrees of freedom                   1    |
| Prob[ChiSqd > value] =               .1889107 |
+-----+
| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
+-----+
| Characteristics in numerator of Prob[Y = 1] |
| Constant | -1.386294361 | 1.1180340 | -1.240 | .2150 |
| GENDER   | 1.791759469 | 1.4433757 | 1.241 | .2145 | .50000000
    
```

THE CHI-SQUARED IS THE DIFFERENCE BETWEEN THE TWO LOG LIKELIHOODS MULTIPLIED BY -2

$CHI-SQUARE = -2 (RESTRICTED LOG LIKELIHOOD - LOG LIKELIHOOD FUNCTION) =$   
 $-2 * [-6.730117 - (-5.867070)] = -2 * (0.863047) = 1.726094$  (small rounding error)

```

+-----+
| Fit Measures for Binomial Choice Model |
| Logit model for variable UNEMP |
+-----+
| Proportions P0= .600000 P1= .400000 |
| N = 10 N0= 6 N1= 4 |
| LogL= -5.867 LogL0= -6.730 |
| Estrella = 1 - (L/L0) ^ (-2L0/n) = .16867 |
+-----+
| Efron | McFadden | Ben./Lerman |
| .16667 | .12824 | .60000 |
| Cramer | Veall/Zim. | Rsqrd ML |
| .16667 | .25656 | .15853 |
+-----+
| Information Akaike I.C. Schwarz I.C. |
| Criteria 1.57341 1.63393 |
+-----+

```

$L0=LogL0$  and  $L=LogL$

```

--> NAMELIST; X=ONE,GENDER$
--> CALC; K=COL(X)$
--> MATRIX; B1=B(1:K)$
--> CREATE; P0=1/(1+EXP(X'B1)); P1=1-P0$
--> create; Res=unemp-P1$
--> create; Ressq=Res*Res$
--> create; Dev=unemp-.4$
--> create; Devsq=Dev*Dev$
--> list; unemp, p1, Res, Ressq, Dev, Devsq$

```

This creates the predicted probabilities ( $P1$  or  $\pi_i$ )  
the residuals ( $RES=UNEMP-P1$  or  $y_i - \pi_i$ )  
the squared residuals ( $RESSQ$ )  
the dependent variable as deviation from its mean  
( $DEV=UNEMP-Mean(UNEMP)$  or  $y_i - Mean(y)$ ) and its  
squared value ( $DEVSQ$ )

Obs.	UNEMP	P1	RES ( $y_i - \pi_i$ )	RESSQ ( $(y_i - \pi_i)^2$ )	DEV ( $y_i - Mean(y)$ )	DEVSQ ( $(y_i - Mean(y))^2$ )
1	0	.60000	-.60000	.36000	-.40000	.16000
2	1	.20000	.80000	.64000	.60000	.36000
3	1	.60000	.40000	.16000	.60000	.36000
4	1	.60000	.40000	.16000	.60000	.36000
5	1	.60000	.40000	.16000	.60000	.36000
6	0	.20000	-.20000	.04000	-.40000	.16000
7	0	.20000	-.20000	.04000	-.40000	.16000
8	0	.20000	-.20000	.04000	-.40000	.16000
9	0	.60000	-.60000	.36000	-.40000	.16000
10	0	.20000	-.20000	.04000	-.40000	.16000
SUM				2.000		2.4

**Efron is a residual based measure like R-square:  $1 - SSReg/TSS$ .**

where  $SSReg = \sum (y_i - \pi_i)^2$   $y_i$  takes 1 if Unemployed, 0 if not,  
 $\pi_i$  is the predicted probability, and  
 $TSS = \sum (y_i - Mean(y))^2$

$$Efron = \frac{\sum (DEVSQ) - \sum (RESSQ)}{\sum (DEVSQ)} = 1 - \frac{\sum (RESSQ)}{\sum (DEVSQ)} = 1 - \frac{2}{2.4} = 1 - .833333 = 0.166667$$

**McFadden and Estrella are log likelihood based measures.**

$L$  is the value of the Log likelihood Function and  
 $L0$  is the value of the Restricted log likelihood

$$McFadden = \frac{(L0 - L)}{L0} = 1 - \frac{L}{L0} = 1 - \frac{(-5.867070)}{(-6.730117)} = 1 - .87176346 = .1282354$$

$$Estrella = 1 - \frac{(L/L0)^{-2L0/n}}{1} = 1 - \left[ \frac{(-5.867070)}{(-6.730117)} \right]^{-2 * (-6.730117)/10} = 1 - .87176346^{1.346} = 1 - .7082387 = 1 - .83133 = .16867$$