

List 09: logit/probit-регрессион

N.V. Artamonov

Содержание

1 Fitting & coefficients' interpretation	1
1.1 approve equation #1 (probit)	1
1.2 approve equation #2 (logit)	2
1.3 labour force equation #1 (probit)	2
1.4 labour force equation #2 (logit)	2
2 z-test	3
2.1 approve equation #1 (probit)	3
2.2 approve equation #2 (logit)	3
2.3 labour force equation #1 (probit)	4
2.4 labour force equation #2 (logit)	4
3 LR-test: overall significance	5
3.1 approve equation #1 (probit)	5
3.2 approve equation #2 (logit)	5
3.3 labour force equation #1 (probit)	6
3.4 labour force equation #2 (logit)	7
4 Wald-test: joint significance	9
4.1 swiss labour force equation #1	9

1 Fitting & coefficients' interpretation

1.1 approve equation #1 (probit)

For the dataset `loanapp` consider probit-regression **approve** on **appinc**, **mortno**, **unem**, **dep**, **male**, **married**, **yjob**, **self**

Specification: $P(\text{approve} = 1) = \Phi(\beta_0 + \beta_1 \text{appinc} + \beta_2 \text{mortno} + \beta_3 \text{unem} + \beta_4 \text{dep} + \beta_5 \text{male} + \beta_6 \text{married} + \beta_7 \text{yjob} + \beta_8 \text{self})$

An alternative specification: $\text{probit}(P(\text{approve} = 1)) = \beta_0 + \beta_1 \text{appinc} + \beta_2 \text{mortno} + \beta_3 \text{unem} + \beta_4 \text{dep} + \beta_5 \text{male} + \beta_6 \text{married} + \beta_7 \text{yjob} + \beta_8 \text{self}$

Fit the model and report coefficients. **Round the answer to 3 decimal places.**

The answer:

(Intercept)	appinc	mortno	unem	dep	male
1.142	-0.001	0.407	-0.031	-0.083	0.020
married	yjob	self			
0.221	-0.001	-0.158			

Give the interpretation of coefficients.

1.2 approve equation #2 (logit)

For the dataset `loanapp` consider logit-regression **approve** on **appinc**, **mortno**, **unem**, **dep**, **male**, **married**, **yjob**, **self**

Specification: $P(\text{approve} = 1) = \Lambda(\beta_0 + \beta_1 \text{appinc} + \beta_2 \text{mortno} + \beta_3 \text{unem} + \beta_4 \text{dep} + \beta_5 \text{male} + \beta_6 \text{married} + \beta_7 \text{yjob} + \beta_8 \text{self})$

An alternative specification: $\text{logit}(P(\text{approve} = 1)) = \beta_0 + \beta_1 \text{appinc} + \beta_2 \text{mortno} + \beta_3 \text{unem} + \beta_4 \text{dep} + \beta_5 \text{male} + \beta_6 \text{married} + \beta_7 \text{yjob} + \beta_8 \text{self}$

Здесь $\text{logit}(P(\text{approve} = 1)) = \log \frac{P(\text{approve}=1)}{1-P(\text{approve}=1)} = \log \frac{P(\text{approve}=1)}{P(\text{approve}=0)}$

Fit the model and report coefficients. **Round the answer to 3 decimal places.**

The answer:

(Intercept)	appinc	mortno	unem	dep	male
1.931	-0.001	0.787	-0.055	-0.161	0.030
married	yjob	self			
0.425	-0.006	-0.280			

Give the interpretation of coefficients..

1.3 labour force equation #1 (probit)

For the dataset `TableF5-1` consider probit-regression **LFP** on **WA**, **WA^2**, **WE**, **KL6**, **K618**, **CIT**, **UN**, **log(FAMINC)**

Specification: $P(LFP = 1) = \Phi(\beta_0 + \beta_1 WA + \beta_2 WA^2 + \beta_3 WE + \beta_4 KL6 + \beta_5 K618 + \beta_6 CIT + \beta_7 UN + \beta_8 \log(FAMINC))$

An alternative specification: $\text{probit}(P(LFP = 1)) = \beta_0 + \beta_1 WA + \beta_2 WA^2 + \beta_3 WE + \beta_4 KL6 + \beta_5 K618 + \beta_6 CIT + \beta_7 UN + \beta_8 \log(FAMINC)$

Fit the model and report coefficients. **Round the answer to 3 decimal places.**

The answer:

(Intercept)	WA	I (WA^2)	WE	KL6	K618
-2.005	0.008	-0.001	0.109	-0.851	-0.063
CIT	UN	log (FAMINC)			
-0.128	-0.011	0.200			

Give the interpretation of coefficients.

1.4 labour force equation #2 (logit)

For the dataset `TableF5-1` consider logit-regression **LFP** on **WA**, **WA^2**, **WE**, **KL6**, **K618**, **CIT**, **UN**, **log(FAMINC)**

Specification: $P(LFP = 1) = \Lambda(\beta_0 + \beta_1 WA + \beta_2 WA^2 + \beta_3 WE + \beta_4 KL6 + \beta_5 K618 + \beta_6 CIT + \beta_7 UN + \beta_8 \log(FAMINC))$

An alternative specification: $\text{logit}(P(LFP = 1)) = \beta_0 + \beta_1 WA + \beta_2 WA^2 + \beta_3 WE + \beta_4 KL6 + \beta_5 K618 + \beta_6 CIT + \beta_7 UN + \beta_8 \log(FAMINC)$

Здесь $\text{logit}(P(LFP = 1)) = \log \frac{P(LFP=1)}{1-P(LFP=1)} = \log \frac{P(LFP=1)}{P(LFP=0)}$

Fit the model and report coefficients. **Round the answer to 3 decimal places.**

The answer:

(Intercept)	WA	I (WA^2)	WE	KL6	K618
-3.241	0.007	-0.001	0.180	-1.414	-0.104
CIT	UN	log (FAMINC)			
-0.217	-0.018	0.333			

Give the interpretation of coefficients..

2 z-test

2.1 approve equation #1 (probit)

For the dataset `loanapp` consider probit-regression **approve** on **appinc, mortno, unem, dep, male, married, yjob, self**

Fit the model and report the output result of z-test

The answer:

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.1418	0.1085	10.5241	<2e-16	***
appinc	-0.0005	0.0004	-1.3564	0.1750	
mortno	0.4071	0.0869	4.6840	<2e-16	***
unem	-0.0308	0.0162	-1.8961	0.0579	.
dep	-0.0828	0.0352	-2.3558	0.0185	*
male	0.0200	0.0998	0.2002	0.8413	
married	0.2208	0.0869	2.5394	0.0111	*
yjob	-0.0007	0.0345	-0.0202	0.9839	
self	-0.1583	0.1073	-1.4751	0.1402	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Significance level is 10%

Evaluate critical value. **Round the answer to 3 decimal places.**

```
[1] 1.645
```

Which coefficients are significant? The answer

```
[1] "(Intercept)" "mortno" "unem" "dep" "married"
```

2.2 approve equation #2 (logit)

For the dataset `loanapp` consider logit-regression **approve** on **appinc, mortno, unem, dep, male, married, yjob, self**

Fit the model and report the output result of z-test

The answer:

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.9315	0.1993	9.6891	<2e-16	***
appinc	-0.0010	0.0007	-1.4717	0.1411	
mortno	0.7868	0.1721	4.5714	<2e-16	***

```

unem          -0.0549      0.0294 -1.8661      0.0620 .
dep           -0.1608      0.0647 -2.4861      0.0129 *
male          0.0300      0.1859  0.1612      0.8719
married       0.4246      0.1624  2.6145      0.0089 **
yjob          -0.0065      0.0651 -0.0993      0.9209
self         -0.2804      0.1967 -1.4257      0.1539
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Significance level is 5%

Evaluate critical value. **Round the answer to 3 decimal places.**

```
[1] 1.96
```

Which coefficients are significant? The answer

```
[1] "(Intercept)" "mortno"      "dep"      "married"
```

2.3 labour force equation #1 (probit)

For the dataset TableF5-1 consider probit-regression **LFP on WA, WA², WE, KL6, K618, CIT, UN, log(FAMINC)**

Fit the model and report the output result of z-test

The answer:

z test of coefficients:

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.0046      1.7039 -1.1765  0.2394
WA           0.0076      0.0701  0.1087  0.9135
I (WA^2)     -0.0005      0.0008 -0.6554  0.5122
WE           0.1088      0.0241  4.5144 <2e-16 ***
KL6          -0.8513      0.1154 -7.3778 <2e-16 ***
K618         -0.0632      0.0417 -1.5157  0.1296
CIT          -0.1277      0.1070 -1.1932  0.2328
UN           -0.0106      0.0157 -0.6771  0.4983
log (FAMINC)  0.1996      0.1049  1.9021  0.0572 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Significance level is 10%

Evaluate critical value. **Round the answer to 3 decimal places.**

```
[1] 1.645
```

Which coefficients are significant? The answer

```
[1] "WE"      "KL6"      "log (FAMINC) "
```

2.4 labour force equation #2 (logit)

For the dataset TableF5-1 consider logit-regression **LFP on WA, WA², WE, KL6, K618, CIT, UN, log(FAMINC)**

Fit the model and report the output result of z-test

The answer:

z test of coefficients:

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.2407      2.8337  -1.1436  0.2528
WA            0.0070      0.1159   0.0602  0.9520
I (WA^2)     -0.0008      0.0013  -0.6061  0.5444
WE           0.1800      0.0404   4.4535 <2e-16 ***
KL6          -1.4138      0.1987  -7.1152 <2e-16 ***
K618         -0.1042      0.0687  -1.5166  0.1294
CIT          -0.2165      0.1765  -1.2267  0.2199
UN           -0.0176      0.0258  -0.6812  0.4957
log (FAMINC)  0.3331      0.1729   1.9272  0.0540 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Significance level is 5%

Evaluate critical value. **Round the answer to 3 decimal places.**

```
[1] 1.96
```

Which coefficients are significant? The answer

```
[1] "WE" "KL6"
```

3 LR-test: overall significance

3.1 approve equation #1 (probit)

For the dataset `loanapp` consider probit-regression **approve on appinc, unem, male, yjob, self**

Fit the model and test overall significance. Significance level is 10%.

Evaluate test statistics. **Round the answer to 3 decimal places.**

The answer:

```
[1] 8.573
```

Evaluate critical value. **Round the answer to 3 decimal places.**

```
[1] 9.236
```

Is the regression significant? The answer

```
[1] "Insignificant"
```

Which coefficients are significant?

```
[1] "(Intercept)" "unem"
```

3.2 approve equation #2 (logit)

For the dataset `loanapp` consider logit-regression **approve on appinc, appinc^2, mortno, unem, dep, male, married, yjob, self**

Fit the model and test overall significance. Significance level is 5%.

Evaluate test statistics. **Round the answer to 3 decimal places.**

The answer:

[1] 48.496

Evaluate critical value. **Round the answer to 3 decimal places.**

[1] 16.919

Is regression significant? The answer

[1] "Significant"

Which coefficients are significant?

[1] "(Intercept)" "I (appinc^2)" "mortno" "dep" "married"

3.3 labour force equation #1 (probit)

For the dataset TableF5-1 consider a collection of probit-regression. Fitting results

=====				
	Зависимая переменная			

	LFP			
	(1)	(2)	(3)	(4)

WA	0.0076 (0.0701)	0.1084* (0.0635)	0.1297** (0.0630)	
I (WA2)	-0.0005 (0.0008)	-0.0014* (0.0007)	-0.0016** (0.0007)	
WE	0.1088*** (0.0241)			
KL6	-0.8513*** (0.1154)			
K618	-0.0632 (0.0417)			
CIT	-0.1277 (0.1070)	-0.1026 (0.1029)	0.0053 (0.0983)	-0.0024 (0.0975)
UN	-0.0106 (0.0157)	-0.0101 (0.0152)	-0.0102 (0.0151)	-0.0115 (0.0150)
log (FAMINC)	0.1996* (0.1049)	0.3621*** (0.0957)		
Constant	-2.0046 (1.7039)	-5.2365*** (1.5600)	-2.1745 (1.3281)	0.2733* (0.1410)

Observations	753	753	753	753
Log Likelihood	-462.3402	-502.2236	-509.6532	-514.5631
Akaike Inf. Crit.	942.6804	1016.4470	1029.3060	1035.1260
=====				

Note:

*p<0.1; **p<0.05; ***p<0.01

For each regression evaluate LR-statistics for overall significance. **Round the answer to 3 decimal places.**

The answer

```
=====
Regression LR.stat
-----
1          105.066
2          25.299
3          10.440
4           0.620
-----
```

For each regression evaluate necessary critical value. Significance level is 10%. **Round the answer to 3 decimal places.**

The answer

```
=====
Regression Critical
-----
1          13.362
2           9.236
3           7.779
4           4.605
-----
```

Which regression is significant?

```
=====
Regression Significance
-----
1          Significant
2          Significant
3          Significant
4          Insignificant
-----
```

3.4 labour force equation #2 (logit)

For the dataset TableF5-1 consider a collection of logit-regressions. fitting results

```
=====
                                Зависимая переменная
                                -----
                                LFP
                                (1)      (2)      (3)      (4)
-----
WA          0.0070
            (0.1159)

I (WA2)     -0.0008
```

	(0.0013)			
WE	0.1800*** (0.0404)	0.2028*** (0.0396)		
KL6	-1.4138*** (0.1987)	-1.0154*** (0.1646)	-0.8645*** (0.1585)	
K618	-0.1042 (0.0687)	0.0509 (0.0594)	0.0288 (0.0581)	-0.0031 (0.0558)
CIT	-0.2165 (0.1765)	-0.2753 (0.1726)	-0.2200 (0.1691)	-0.0043 (0.1564)
UN	-0.0176 (0.0258)	-0.0287 (0.0254)	-0.0196 (0.0248)	-0.0185 (0.0240)
log (FAMINC)	0.3331* (0.1729)	0.2808* (0.1683)	0.5829*** (0.1580)	
Constant	-3.2407 (2.8337)	-4.3882*** (1.5718)	-5.0277*** (1.5542)	0.4420* (0.2383)

Observations	753	753	753	753
Log Likelihood	-462.2363	-475.6736	-489.7908	-514.5605
Akaike Inf. Crit.	942.4726	965.3473	991.5816	1037.1210

Note: *p<0.1; **p<0.05; ***p<0.01

For each regression evaluate LR-statistics for overall significance. **Round the answer to 3 decimal places.**

The answer

Regression LR.stat	
1	105.274
2	78.399
3	50.165
4	0.625

For each regression evaluate necessary critical value. Significance level is 5%. **Round the answer to 3 decimal places.**

The answer

Regression Critical	
1	15.507
2	12.592
3	11.070
4	7.815

Which regression is significant?

```
=====
Regression Significance
-----
1          Significant
2          Significant
3          Significant
4          Insignificant
-----
```

4 Wald-test: joint significance

4.1 swiss labour force equation #1

For the dataset `SwissLabour` consider logit-regression **participation on income, income², age, age², youngkids, oldkids, foreign**

Fitting result

MODEL INFO:

Observations: 872

Dependent Variable: `as.numeric(participation) - 1`

Type: Generalized linear model

Family: binomial

Link function: logit

MODEL FIT:

$\chi^2(7) = 185.1659$, $p = 0.0000$

Pseudo-R² (Cragg-Uhler) = 0.2556

Pseudo-R² (McFadden) = 0.1539

AIC = 1034.0575, BIC = 1072.2238

Standard errors:MLE

	Est.	S.E.	z val.	p	VIF
(Intercept)	-9.4763	17.2451	-0.5495	0.5827	
income	1.8753	3.2660	0.5742	0.5658	276.5677
I(income^2)	-0.1377	0.1552	-0.8875	0.3748	276.4963
age	3.4025	0.6866	4.9553	0.0000	83.0214
I(age^2)	-0.4846	0.0851	-5.6916	0.0000	83.3744
youngkids	-1.1813	0.1723	-6.8578	0.0000	1.5869
oldkids	-0.2471	0.0843	-2.9321	0.0034	1.4726
foreignyes	1.0728	0.1870	5.7371	0.0000	1.0847

Fit the model and test the following hypothesis with Wald test (use χ^2 -statistics). Significance level is 5%.

4.1.1 Hypothesis 1

Test the significance of income, i.e. test the hypothesis $H_0 : \beta_{income} = \beta_{income^2} = 0$.

Evaluate test statistics and its P-value.

```
=====
Chisq  Pr(> Chisq)
-----
24.441    0.00000
-----
```

Evaluate critical value. **Round the answer to 3 decimal places.**

```
[1] 5.991
```

Conclusion

```
[1] "Significant"
```

4.1.2 Hypothesis 2

Test the significance of the number of kids, i.e. the hypothesis $H_0 : \beta_{youngkids} = \beta_{oldkids} = 0$.

Evaluate test statistics and its P-value.

```
=====
Chisq  Pr(> Chisq)
-----
48.420      0
-----
```

Evaluate critical value. **Round the answer to 3 decimal places.**

```
[1] 5.991
```

Conclusion

```
[1] "Significant"
```

4.1.3 Hypothesis 3

Test the significance of age, i.e. the hypothesis $H_0 : \beta_{age} = \beta_{age^2} = 0$.

Evaluate test statistics and its P-value.

```
=====
Chisq  Pr(> Chisq)
-----
58.911      0
-----
```

Evaluate critical value. **Round the answer to 3 decimal places.**

```
[1] 5.991
```

Conclusion

```
[1] "Significant"
```