

Задачи по Эконометрике-2: logit/probit-модель

Н.В. Артамонов (МГИМО МИД России)

Содержание

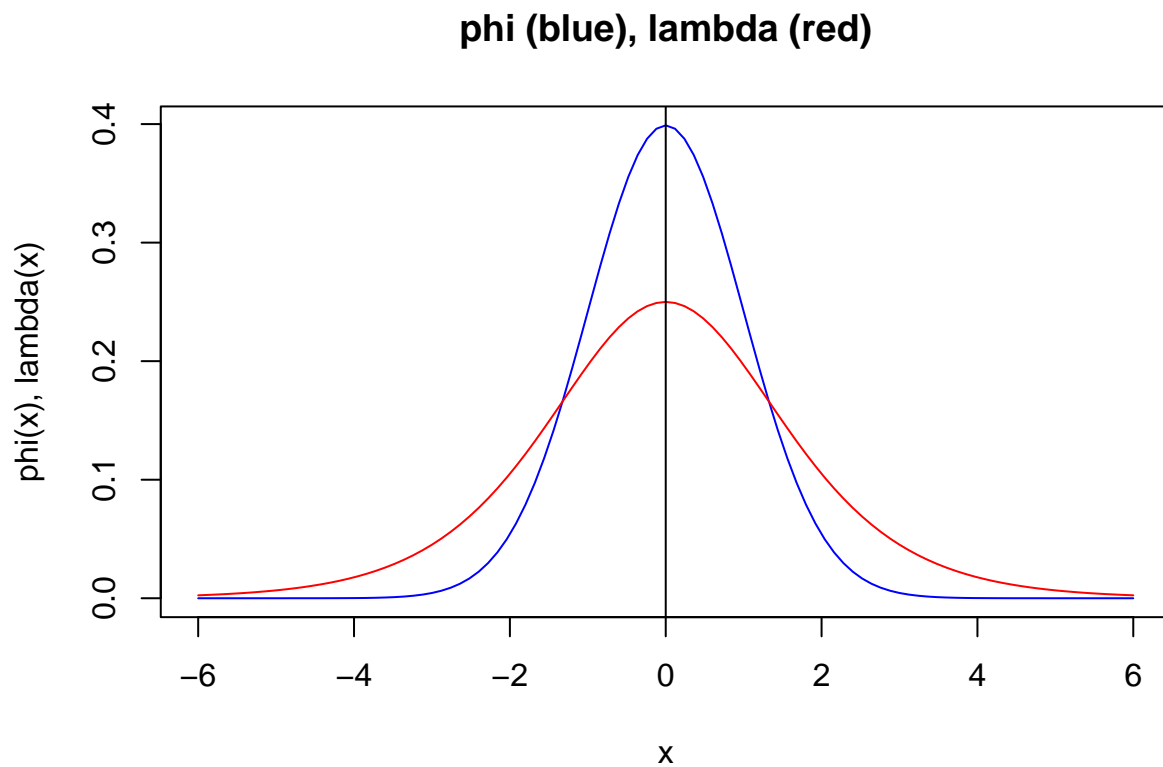
1	Нормальное и логистическое распределение	1
2	Оценивание и интерпретация коэффициентов	3
2.1	approve equation #1 (probit)	3
2.2	approve equation #2 (logit)	3
2.3	labour force equation #1 (probit)	4
2.4	labour force equation #2 (logit)	4
3	z-test	5
3.1	approve equation #1 (probit)	5
3.2	approve equation #2 (logit)	5
3.3	labour force equation #1 (probit)	6
3.4	labour force equation #2 (logit)	6
4	LR-тест: значимость регрессии	7
4.1	approve equation #1 (probit)	7
4.2	approve equation #2 (logit)	8
4.3	labour force equation #1 (probit)	8
4.4	labour force equation #2 (logit)	10
5	LR-тест: совместная значимость	11
5.1	labour force equation #1 (probit)	11
5.2	approve equation #1 (logit)	12
6	Тест Вальда: совместная значимость	13
6.1	swiss labour force equation #1	13

1 Нормальное и логистическое распределение

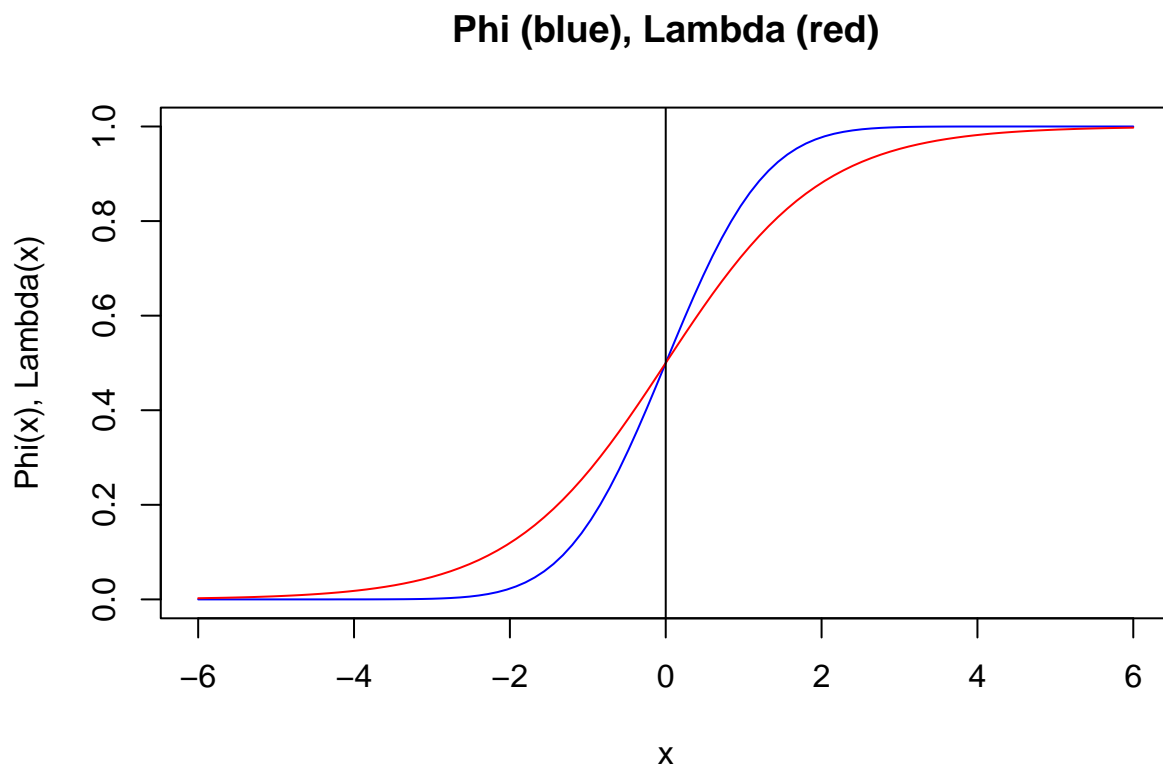
Плотность стандартного $N(0,1)$ гауссова распределения $\phi(t) = \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)$ ($t \in \mathbb{R}$). Функция стандартного нормального распределения $\Phi(x) = \int_{-\infty}^x \phi(t)dt$ ($x \in \mathbb{R}$).

Функция логистического распределения $\Lambda(x) = \frac{\exp(x)}{1+\exp(x)}$ ($x \in \mathbb{R}$). Плотность логистического распределения $\lambda(x) = \Lambda'(x) = \frac{\exp(x)}{(1+\exp(x))^2}$

Плотности на одном графике

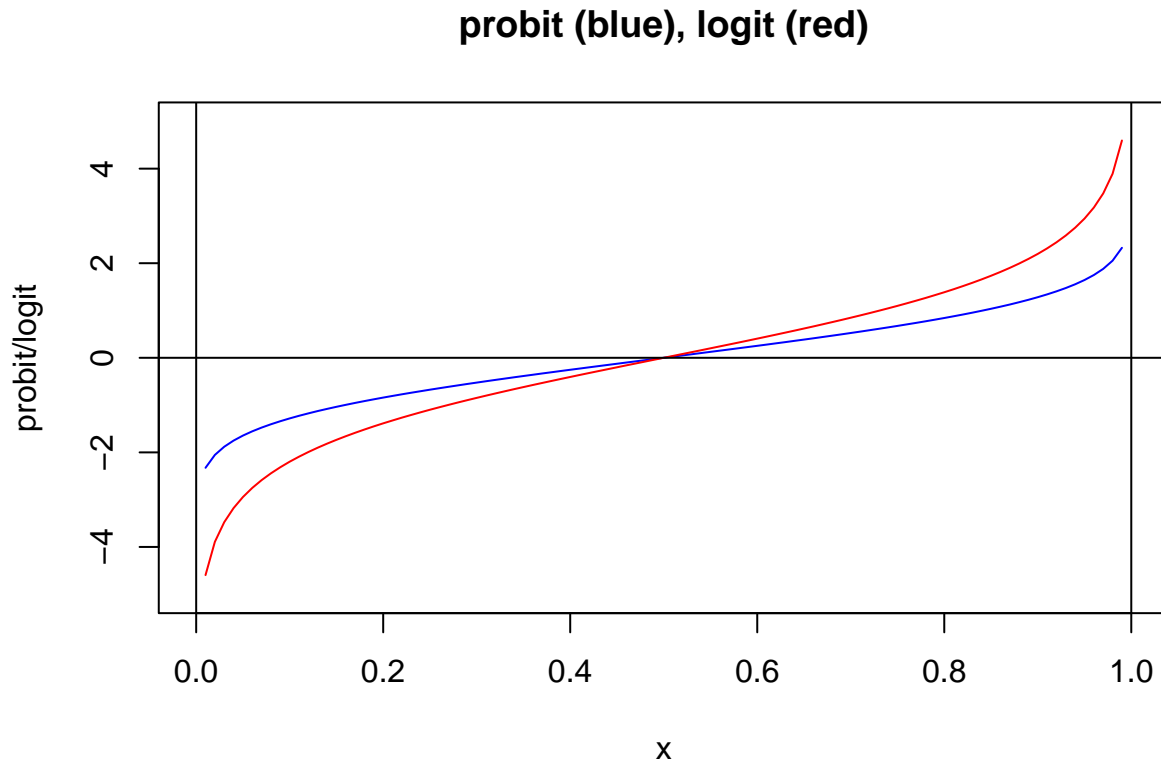


Функции распределения на одном графике



Обратные функции распределение $\text{logit}(p) = \Lambda^{-1}(p) = \log \frac{p}{1-p}$ и $\text{probit}(p) = \Phi^{-1}(p)$, $0 < p < 1$

Их графики



2 Оценивание и интерпретация коэффициентов

2.1 approve equation #1 (probit)

Для датасета `loanapp` рассмотрим probit-регрессию `approve` на `appinc`, `mortno`, `unem`, `dep`, `male`, `married`, `yjob`, `self`

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

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

Оцените модель на данных и укажите коэффициенты подогнанной модели. **Ответ округлите до 3-х десятичных знаков.**

Ответ:

(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			

Дайте интерпретацию коэффициентам модели.

2.2 approve equation #2 (logit)

Для датасета `loanapp` рассмотрим logit-регрессию `approve` на `appinc`, `mortno`, `unem`, `dep`, `male`, `married`, `yjob`, `self`

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

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

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

Оцените модель на данных и укажите коэффициенты подогнанной модели. **Ответ округлите до 3-х десятичных знаков.**

Ответ:

(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			

Дайте интерпретацию коэффициентам модели.

2.3 labour force equation #1 (probit)

Для датасета mroz_Greene рассмотрим probit-регрессию **LFP** на **WA**, **WA^2**, **WE**, **KL6**, **K618**, **CIT**, **UN**, **log(FAMINC)**

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

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

Оцените модель на данных и укажите коэффициенты подогнанной модели. **Ответ округлите до 3-х десятичных знаков.**

Ответ:

(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			

Дайте интерпретацию коэффициентам модели.

2.4 labour force equation #2 (logit)

Для датасета mroz_Greene рассмотрим logit-регрессию **LFP** на **WA**, **WA^2**, **WE**, **KL6**, **K618**, **CIT**, **UN**, **log(FAMINC)**

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

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

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

Оцените модель на данных и укажите коэффициенты подогнанной модели. **Ответ округлите до 3-х десятичных знаков.**

Ответ:

(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			

Дайте интерпретацию коэффициентам модели.

3 z-test

3.1 approve equation #1 (probit)

Для датасета `loanapp` рассмотрим probit-регрессию `approve` на `appinc`, `mortno`, `unem`, `dep`, `male`, `married`, `yjob`, `self`

Подгоните модель на данных и приведите результаты z-теста

Ответ:

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

Модель была подогнана по 1971 наблюдениям. Уровень значимости 10%

Вычислите необходимое критическое значение z-теста. **Ответ округлите до 3-х десятичных знаков.**

[1] 1.645

Какие коэффициенты значимы? Ответ

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

3.2 approve equation #2 (logit)

Для датасета `loanapp` рассмотрим logit-регрессию `approve` на `appinc`, `mortno`, `unem`, `dep`, `male`, `married`, `yjob`, `self`

Подгоните модель на данных и приведите результаты z-теста

Ответ:

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
```

Модель была подогнана по 1971 наблюдениям. Уровень значимости 5%

Вычислите необходимое критическое значение z-теста. **Ответ округлите до 3-х десятичных знаков.**

```
[1] 1.96
```

Какие коэффициенты значимы? Ответ

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

3.3 labour force equation #1 (probit)

Для датасета mroz_Greene рассмотрим probit-регрессию **LFP** на **WA, WA^2, WE, KL6, K618, CIT, UN, log(FAMINC)**

Подгоните модель на данных и приведите результаты z-теста

Ответ:

```
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
```

Модель была подогнана по 753 наблюдениям. Уровень значимости 10%

Вычислите необходимое критическое значение z-теста. **Ответ округлите до 3-х десятичных знаков.**

```
[1] 1.645
```

Какие коэффициенты значимы? Ответ

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

3.4 labour force equation #2 (logit)

Для датасета mroz_Greene рассмотрим logit-регрессию **LFP** на **WA, WA^2, WE, KL6, K618, CIT, UN, log(FAMINC)**

Подгоните модель на данных и приведите результаты z-теста

Ответ:

```
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

Модель была подогнана по 753 наблюдениям. Уровень значимости 5%

Вычислите необходимое критическое значение z-теста. **Ответ округлите до 3-х десятичных знаков.**

```
[1] 1.96
```

Какие коэффициенты значимы? Ответ

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

4 LR-тест: значимость регрессии

4.1 approve equation #1 (probit)

Для датасета `loanapp` рассмотрим probit-регрессию `approve` на `appinc`, `unem`, `male`, `yjob`, `self`

Подгоните модель на данных и укажите:

- число наблюдений, на которых была подогнана модель
- логарфим функции правдоподобия для модели $\log L = -734.0891$,
- логарфим функции правдоподобия для регрессии без объясняющих переменных (только на константу)

Ответ округлите до 3-х десятичных знаков.

Ответ:

```
[1] "1974 -734.0891 -738.3755"
```

Тестируется значимость регрессии, т.е. гипотеза $H_0 : \beta_{appinc} = \beta_{unem} = \beta_{male} = \beta_{yjob} = \beta_{self} = 0$. Уровень значимости 10%.

Вычислите тестовую статистику. **Ответ округлите до 3-х десятичных знаков.**

Ответ:

```
[1] 8.573
```

Вычислите критическое значение. **Ответ округлите до 3-х десятичных знаков.**

```
[1] 9.236
```

Значима ли регрессия? Ответ

```
[1] "Незначима"
```

Какие коэффициенты значимы?

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

4.2 approve equation #2 (logit)

Для датасета `loanapp` рассмотрим logit-регрессию `approve` на `appinc`, `appinc^2`, `mortno`, `unem`, `dep`, `male`, `married`, `yjob`, `self`

Подгоните модель на данных и укажите:

- число наблюдений, на которых была подогнана модель
- логарфим функции правдоподобия для модели $\log L = -713.7313$,
- логарфим функции правдоподобия для регрессии без объясняющих переменных (только на константу)

Ответ округлите до 3-х десятичных знаков.

Ответ:

```
[1] "1971 -713.7313 -737.9793"
```

Тестируется значимость регрессии, т.е. гипотеза $H_0 : \beta_{appinc} = \beta_{unem} = \beta_{male} = \beta_{yjob} = \beta_{self} = 0$. Уровень значимости 5%.

Вычислите тестовую статистику. **Ответ округлите до 3-х десятичных знаков.**

Ответ:

```
[1] 48.496
```

Вычислите критическое значение. **Ответ округлите до 3-х десятичных знаков.**

```
[1] 16.919
```

Значима ли регрессия? Ответ

```
[1] "Значима"
```

Какие коэффициенты значимы?

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

4.3 labour force equation #1 (probit)

Для датасета `mroz_Greene` рассмотрим несколько probit-регрессий. Результаты оценивания

	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

Для каждой регрессии вычислите LR-статистику для тестирования её значимости. **Ответ округлите до 3-х десятичных знаков.**

Ответ

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

Для каждой регрессии вычислите необходимое критическое значение Уровень значимости 10%. **Ответ округлите до 3-х десятичных знаков.**

Ответ

```
=====
Регрессия Критическое
-----
1      13.362
2       9.236
3       7.779
4       4.605
-----
```

Какая из регрессий значима?

```
=====
Регрессия Значимость
-----
1      Значима
```

Для датасета mroz Greene рассмотрим несколько logit-регрессий. Результаты оценивания

=====

Для каждой регрессии вычислите LR-статистику для тестирования её значимости. **Ответ округлите до 3-х десятичных знаков.**

=====	
Регрессия	LR.stat

1	105.274
2	78.399
3	50.165
4	0.625

Для каждой регрессии вычислите необходимое критическое значение Уровень значимости 5%. **Ответ округлите до 3-х десятичных знаков.**

Ответ

=====	
Регрессия	Критическое

1	15.507
2	12.592
3	11.070
4	7.815

Какая из регрессий значима?

=====	
Регрессия	Значимость

1	Значима
2	Значима
3	Значима
4	Незначима

5 LR-тест: совместная значимость

5.1 labour force equation #1 (probit)

Для датасета mroz_Greene рассмотрим probit-регрессию LFP на WA6 WA^2, WE, KL6, K618, CIT, UN, log(FAMINC).

Оцените регрессию на данных и тестируйте следующие гипотезы, используя LR-статистику. Уровень значимости 10%.

5.1.1 Гипотеза 1

Тестируется значимость влияния возраста, т.е. гипотеза $H_0 : \beta_{WA} = \beta_{WA^2} = 0$ по LR-тесту.

Вычислите соответствующую тестовую статистику. **Ответ округлите до 3-х десятичных знаков.**

[1] 26.718

Вычислите необходимое критическое значение. **Ответ округлите до 3-х десятичных знаков.**

[1] 4.605

Значимо ли влияние возраста? Ответ

[1] "Значимо"

5.1.2 Гипотеза 2

Тестируется совместная значимость влияния **K618, CIT, UN**, т.е. гипотеза $H_0 : \beta_{K618} = \beta_{CIT} = \beta_{UN} = 0$ по LR-тесту.

Вычислите соответствующую тестовую статистику. **Ответ округлите до 3-х десятичных знаков.**

[1] 4.712

Вычислите необходимое критическое значение. **Ответ округлите до 3-х десятичных знаков.**

[1] 6.251

Значимо ли влияние переменных? Ответ

[1] "Незначимо"

5.1.3 Гипотеза 3

Тестируется совместная значимость влияния **K618, CIT, UN, FAMINC**, т.е. гипотеза $H_0 : \beta_{K618} = \beta_{CIT} = \beta_{UN} = \beta_{\log(FAMINC)} = 0$ по LR-тесту.

Вычислите соответствующую тестовую статистику. **Ответ округлите до 3-х десятичных знаков.**

[1] 7.301

Вычислите необходимое критическое значение. **Ответ округлите до 3-х десятичных знаков.**

[1] 7.779

Значимо ли влияние переменных? Ответ

[1] "Незначимо"

5.2 approve equation #1 (logit)

Для датасета `loanapp` рассмотрим probit-регрессию **approve** на **appinc, appinc^2, mortno, unem, dep, male, married, yjob, self**

Оцените регрессию на данных и тестируйте следующие гипотезы, используя LR-статистику. Уровень значимости 1%.

5.2.1 Гипотеза 1

Тестируется значимость влияния дохода, т.е. гипотеза $H_0 : \beta_{appinc} = \beta_{appinc^2} = 0$ по LR-тесту.

Вычислите соответствующую тестовую статистику. **Ответ округлите до 3-х десятичных знаков.**

[1] 7.213

Вычислите необходимое критическое значение. **Ответ округлите до 3-х десятичных знаков.**

[1] 9.21

Значимо ли влияние возраста? Ответ

[1] "Незначимо"

5.2.2 Гипотеза 2

Тестируется совместная значимость влияния **male, yjob, self**, т.е. гипотеза $H_0 : \beta_{male} = \beta_{yjob} = \beta_{self} = 0$ по LR-тесту.

Вычислите соответствующую тестовую статистику. **Ответ округлите до 3-х десятичных знаков.**

[1] 3.229

Вычислите необходимое критическое значение. **Ответ округлите до 3-х десятичных знаков.**

[1] 11.345

Значимо ли влияние переменных? Ответ

[1] "Незначимо"

5.2.3 Гипотеза 3

Тестируется совместная значимость влияния **male, yjob, self, unem, married**, т.е. гипотеза $H_0 : \beta_{male} = \beta_{yjob} = \beta_{self} = \beta_{unem} = \beta_{married} = 0$ по LR-тесту

Вычислите соответствующую тестовую статистику. **Ответ округлите до 3-х десятичных знаков.**

[1] 13.711

Вычислите необходимое критическое значение. **Ответ округлите до 3-х десятичных знаков.**

[1] 15.086

Значимо ли влияние переменных? Ответ

[1] "Незначимо"

6 Тест Вальда: совместная значимость

6.1 swiss labour force equation #1

Для датасета SwissLabour рассмотрим логит-регрессию **participation** на **income, income^2, age, age^2, youngkids, oldkids, foreign**

Результаты оценивания

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- R^2 (Cragg-Uhler) = 0.2556

Pseudo- R^2 (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

Оцените регрессию на данных и тестируйте следующие гипотезы, используя χ^2 -статистику Вальда. Уровень значимости 5%.

6.1.1 Гипотеза 1

Тестируется значимость влияния дохода, т.е. гипотеза $H_0 : \beta_{income} = \beta_{income^2} = 0$.

Вычислите статистику теста Вальда и Р-значение.

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

Вычислите критическое значение. Ответ округлите до 3-х десятичных знаков.

```
[1] 5.991
```

Значимо ли влияние дохода? Ответ

```
[1] "Значимо"
```

6.1.2 Гипотеза 2

Тестируется значимость влияния числа детей, т.е. гипотеза $H_0 : \beta_{youngkids} = \beta_{oldkids} = 0$. Вычислите статистику теста Вальда и Р-значение.

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

Вычислите критическое значение. Ответ округлите до 3-х десятичных знаков.

```
[1] 5.991
```

Значимо ли влияние числе детей? Ответ

```
[1] "Значимо"
```

6.1.3 Гипотеза 3

Тестируется значимость влияния возраста, т.е. гипотеза $H_0 : \beta_{age} = \beta_{age^2} = 0$. Вычислите статистику теста Вальда и Р-значение.

```
=====
Chisq  Pr(> Chisq)
```

58.911 0

Вычислите критическое значение. **Ответ округлите до 3-х десятичных знаков.**

[1] 5.991

Значимо ли влияние возраста? Ответ

[1] "Значимо"