NATIONAL RESEARCH UNIVERSITY

Лекция по эконометрике № 5, 3 модуль

Модели бинарного выбора

Демидова

Ольга Анатольевна

https://www.hse.ru/staff/demidova_olga

E-mail:demidova@hse.ru

11.02.2022

План лекции

1) Модели с бинарной зависимой переменной
2) Модель линейной вероятности, ее недостатки
3) Логит и Пробит модели, их оценивание
4) Интерпретация результатов оценивания моделей с бинарными зависимыми переменными
5) Показатели качества оценки моделей бинарного выбора

Модели с бинарной зависимой переменной

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + \varepsilon_i, i = 1, \ldots, n,$$

$$Y_i = 1 \text{ или } 0$$

Примеры бинарных переменных:

После школы пойти учиться или работать?

Купить квартиру или арендовать?

Участвовать в выборах или нет?

И т.д. и т.п.

Модели с бинарной зависимой переменной

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \varepsilon_i, i = 1, \dots, n,$$
$$Y_i = E(Y_i) + \varepsilon_i$$

Найдем математическое ожидание Y_i.

$p_i = p(Y_i = 1), p(Y_i = 0) = 1 - p_i,$

$E(Y_i) = 1 \times p_i + 0 \times (1 - p_i) = p_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki},$

$p_i = p(Y_i = 1) = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}$

Модель линейной вероятности

$$p_i = p(Y_i = 1) = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}$$

Если мы будем оценивать модель с качественной зависимой переменной, как и ранее, с помощью МНК, мы получим указанную выше модель, называемую моделью линейной вероятности.

Однако линейная вероятностная модель имеет ряд серьезных недостатков:

- **Одним из главных недостатков линейной вероятностной модели является следующий : оцененные значения вероятности могут оказаться больше 1 или меньше 0.**

Модель линейной вероятности

$$p_i = p(Y_i = 1) = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}$$

Однако линейная вероятностная модель имеет ряд серьезных недостатков:

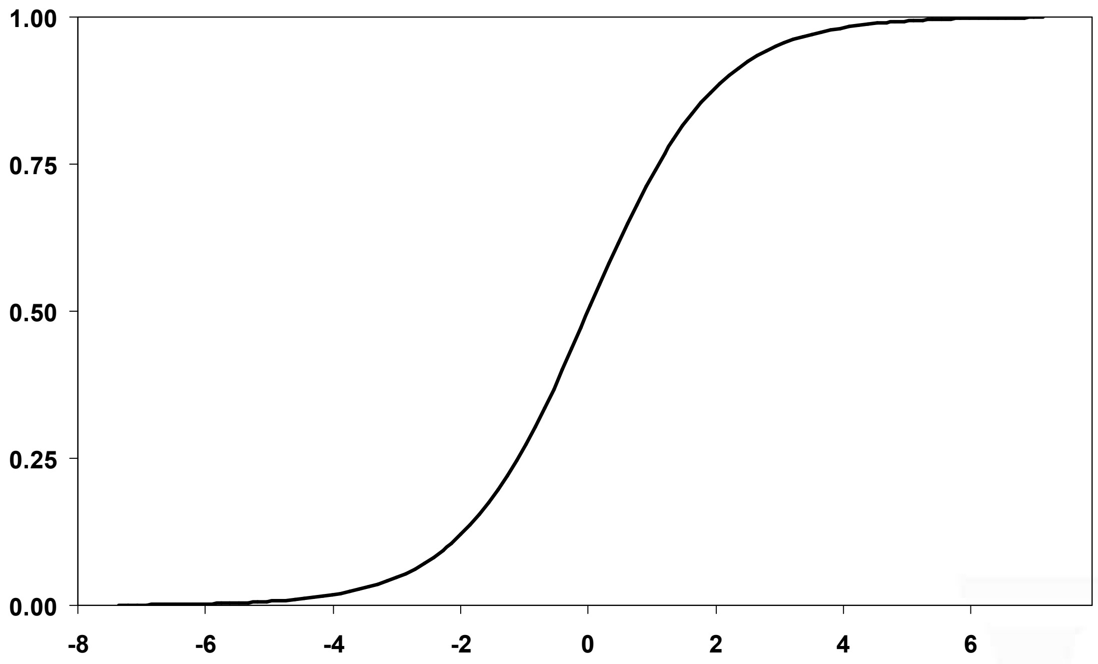
- **Распределение случайного члена не является нормальным,**
- **Можно показать, что дисперсия случайного члена зависит от X, таким образом, имеет место проблема гетероскедастичности.**

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + \varepsilon_i, i = 1, \ldots, n,$$

$$p(Y_i = 1) = P(\varepsilon_i = 1 - \beta_0 - \beta_1 X_{1i} - \cdots - \beta_k X_{ki}) = p_i,$$

$$p(Y_i = 0) = P(\varepsilon_i = 0 - \beta_0 - \beta_1 X_{1i} - \cdots - \beta_k X_{ki}) = 1 - p_i$$

Модели бинарного выбора



Главным недостатком модели линейной вероятности являлась возможность для оцененных значений зависимой переменной принимать значение вне интервала (0, 1).

Обычным способом решения этой проблемы является предположение о том, что вероятность является S – образной функцией от переменной *Z*, *F*(*Z*) принимает значения на интервале (0, 1), где *Z* является линейной функцией от объясняющих переменных.

Модели бинарного выбора

Другой способ получения модели:

$P(Y_i = 1) = \mathrm{F}(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki})$, $i=1,\ldots,n$, (*)

Предположим, что существует латентная переменная Y_i^*, связанная с переменной X обычным регрессионным уравнением:

$Y_i^* = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + \varepsilon_i$, $i=1,\ldots,n$,

где возмущения ε_i, $i=1,\ldots,n$, **независимы и одинаково распределены,**

$\mathrm{E}(\varepsilon_\mathrm{i}) = 0$, $\mathrm{var}(\varepsilon_\mathrm{i}) = \sigma^2$, **и F – функция распределения нормированных возмущений с симметричной функцией плотности.**

$Y_i = 1$, *если* $Y_i^* \geq 0$, $i=1,\ldots,n$,

$Y_i = 0$, *если* $Y_i^* < 0$, $i=1,\ldots,n$

Модели бинарного выбора

$P(Y_i = 1) = F(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki})$ (*),

$Y_i^* = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + \varepsilon_i$, $i=1,\ldots,n$,

$Y_i = 1$, *если* $Y_i^* \geq 0$, $i=1,\ldots,n$,

$Y_i = 0$, *если* $Y_i^* < 0$, $i=1,\ldots,n$

Тогда $P(Y_i = 1) = P(Y_i^* \geq 0) = P(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + \varepsilon_i \geq 0) =$

$= P(\varepsilon_i \geq -\beta_0 - \beta_1 X_{1i} - \ldots - \beta_k X_{ki}) = P(\varepsilon_i \leq \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki}) =$

$= P(\varepsilon_i/\sigma \leq (\beta_0/\sigma) + (\beta_1/\sigma) X_{1i} + \ldots + (\beta_k/\sigma) X_{ki}) =$

$= F((\beta_0/\sigma) + (\beta_1/\sigma) X_{1i} + \ldots + (\beta_k/\sigma) X_{ki})$

что с точностью до нормировки совпадает с (*).

Оценка параметров модели бинарного выбора

$L(\beta) = \prod_{Y_i=1} P(Y_i = 1) \prod_{Y_i=0} P(Y_i = 0) =$

$= \prod_{Y_i=1} F(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}) \prod_{Y_i=0} \left(1 - F(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})\right) =$

$= \prod_{i=1}^{n} [F(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})]^{Y_i} [1 - F(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})]^{1-Y_i} =$

$$l(\beta) = \ln L(\beta) = \sum_{i=1}^{n} [Y_i \ln F(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}) +$$

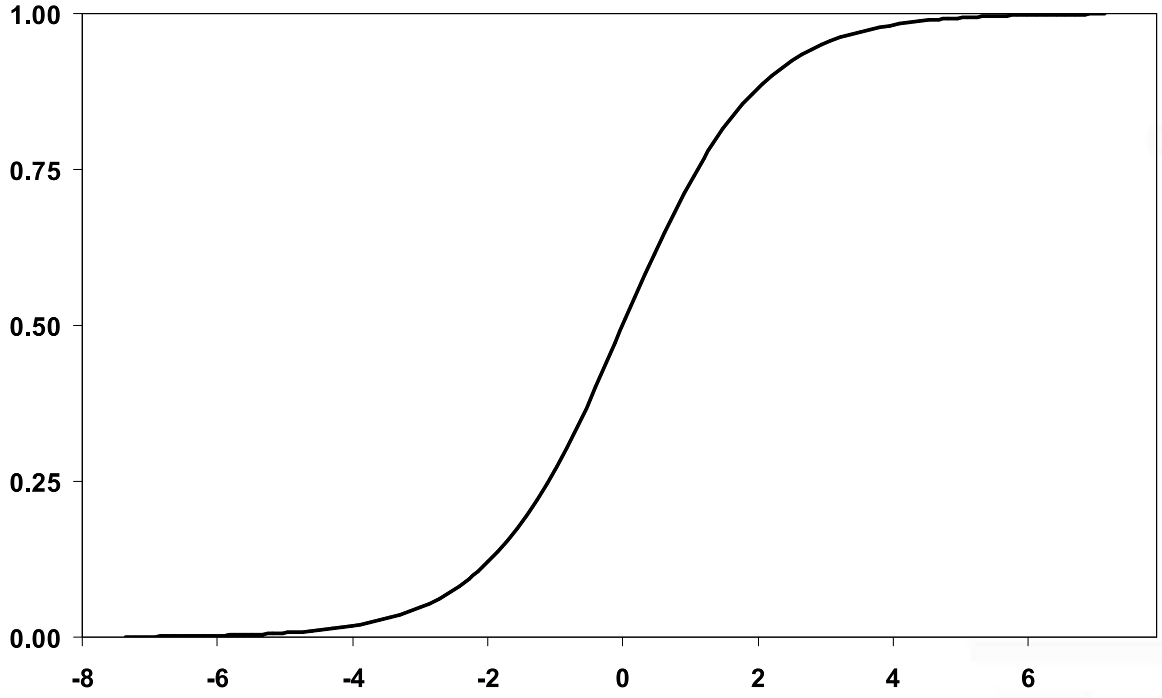
$$+(1 - Y_i) \ln(1 - F(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}))]$$

Необходимые условия экстремума первого порядка

$$\frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} X_{ji} \left[\begin{array}{c} \dfrac{f(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})}{F(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})} Y_i + \\ +(1 - Y_i) \dfrac{(-f(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}))}{1 - F(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})} \end{array} \right] = 0,$$

$j = 0, \ldots, k$

Логит-модель



$$P(Y=1) = F(Z) = \Lambda(Z) = \frac{1}{1+e^{-Z}},$$

Если функция F является логистической (формула для этой функции приведена выше), то соответствующая модель называется логит - моделью.

$f(z) = F'(Z) = \frac{e^{-Z}}{(1+e^{-Z})^2}$ **- функция плотности для логит-модели.**

Оценка параметров логит-модели

$$l(\beta) = \ln L(\beta) = \sum_{i=1}^{n} [Y_i \ln F(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}) +$$

$$+(1 - Y_i)\ln(1 - F(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}))]$$

Для логит-модели необходимые условия экстремума первого порядка

$$\frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} X_{ji} \left[\begin{array}{c} \dfrac{f(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})}{F(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})} Y_i + \\ +(1 - Y_i) \dfrac{(-f(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}))}{1 - F(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})} \end{array} \right] = 0$$

могут быть сведены к виду: $\sum_{i=1}^{n} [Y_i - \Lambda(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})] X_{ji} = 0$,

Матрица Гессе для логит-модели является отрицательно определенной (детали см. Green, издание 7, с.691-692), поэтому найденное решение является максимумом.

Пробит-модель

Если функция плотности является функцией плотности стандартного нормального распределения

$$f(z)=\frac{\mathbf{1}}{\sqrt{\mathbf{2\pi}}}\,exp\left\{-\frac{1}{2}z^2\right\},$$

$$F(Z)=\frac{\mathbf{1}}{\sqrt{\mathbf{2\pi}}}\int_{-\infty}^{\boldsymbol{z}} exp\left\{-\frac{1}{2}x^2\right\}dx,$$

То $P(Y{=}1) = F(Z) = \frac{\mathbf{1}}{\sqrt{\mathbf{2\pi}}}\int_{-\infty}^{\boldsymbol{z}} exp\left\{-\frac{1}{2}x^2\right\}dx$ называется пробит-моделью.

Матрица Гессе для пробит-модели является отрицательно определенной (детали см. Green, издание 7), поэтому найденное решение

$$\frac{\partial l(\beta)}{\partial \beta_j}=\sum_{i=1}^{n}X_{ji}\left[\begin{array}{c}\frac{f(\beta_0+\beta_1X_{1i}+\cdots+\beta_kX_{ki})}{F(\beta_0+\beta_1X_{1i}+\cdots+\beta_kX_{ki})}Y_i+\\ +(1-Y_i)\frac{(-f(\beta_0+\beta_1X_{1i}+\cdots+\beta_kX_{ki}))}{1-F(\beta_0+\beta_1X_{1i}+\cdots+\beta_kX_{ki})}\end{array}\right]=0$$

является максимумом.

Пример использования логит-модели

Пример использования логит – модели для оценки вероятности окончания средней школы. В качестве объясняющей выбрана переменная *ASVABC* (обобщенные результаты тестов).

GRAD = 1, если респондент закончил школу и 0 иначе.

```
. logit GRAD ASVABC

Iteration 0:   log likelihood = -118.67769
Iteration 1:   log likelihood = -104.45292
Iteration 2:   log likelihood = -97.135677
Iteration 3:   log likelihood = -96.887294
Iteration 4:   log likelihood = -96.886017

Logit estimates                                   Number of obs   =        540
                                                  LR chi2(1)      =      43.58
                                                  Prob > chi2     =     0.0000
Log likelihood = -96.886017                       Pseudo R2       =     0.1836

------------------------------------------------------------------------------
        GRAD |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ASVABC |   .1313626    .022428     5.86   0.000      .0874045    .1753206
       _cons |  -3.240218   .9444844    -3.43   0.001     -5.091373   -1.389063
------------------------------------------------------------------------------
```

Пример использования логит-модели

```
. logit GRAD ASVABC

Iteration 0:   log likelihood = -118.67769
Iteration 1:   log likelihood = -104.45292
Iteration 2:   log likelihood = -97.135677
Iteration 3:   log likelihood = -96.887294
Iteration 4:   log likelihood = -96.886017

Logit estimates                                   Number of obs   =        540
                                                  LR chi2(1)      =      43.58
                                                  Prob > chi2     =     0.0000
Log likelihood = -96.886017                       Pseudo R2       =     0.1836

------------------------------------------------------------------------------
        GRAD |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ASVABC |   .1313626    .022428     5.86   0.000     .0874045    .1753206
       _cons |  -3.240218   .9444844    -3.43   0.001    -5.091373   -1.389063
------------------------------------------------------------------------------
```

Для проверки значимости коэффициентов используется z – статистика, поскольку оценки МП имеют (асимптотически) нормальное распределение.
LR статистика используется для проверки совместной значимости коэффициентов при выбранных факторах (всех, кроме константы).

Пример использования пробит-модели

```
. probit GRAD ASVABC SM SF MALE

Iteration 0:   log likelihood = -118.67769
Iteration 1:   log likelihood = -98.195303
Iteration 2:   log likelihood = -96.666096
Iteration 3:   log likelihood = -96.624979
Iteration 4:   log likelihood = -96.624926

Probit estimates                                  Number of obs   =        540
                                                  LR chi2(4)      =      44.11
                                                  Prob > chi2     =     0.0000
Log likelihood = -96.624926                       Pseudo R2       =     0.1858

------------------------------------------------------------------------------
        GRAD |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ASVABC |   .0648442   .0120378     5.39   0.000     .0412505    .0884379
          SM |  -.0081163   .0440399    -0.18   0.854     -.094433    .0782004
          SF |   .0056041   .0359557     0.16   0.876    -.0648677    .0760759
        MALE |   .0630588   .1988279     0.32   0.751    -.3266368    .4527544
       _cons |  -1.450787   .5470608    -2.65   0.008    -2.523006   -.3785673
------------------------------------------------------------------------------
```

Интерпретация результатов оценки моделей бинарного выбора

$P(Y=1) = \mathrm{F}(\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k)$,

$\frac{\partial p}{\partial X_j} = \frac{dp}{dZ}\frac{\partial Z}{\partial X_j} = f(Z)\beta_j$ - предельный эффект непрерывной объясняющей переменной X_j, зависит от точки, в которой этот эффект рассчитывается.

$\frac{\partial p}{\partial X_j} = \frac{dp}{dZ}\frac{\partial Z}{\partial X_j} = f(Z)\beta_j = \frac{e^{-Z}}{(1+e^{-Z})^2}\beta_j$ - для логит-модели

$\frac{\partial p}{\partial X_j} = \frac{dp}{dZ}\frac{\partial Z}{\partial X_j} = f(Z)\beta_j = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2}\right)\beta_j$ - для пробит-модели

$\frac{\partial p}{\partial X_j} = p(X_1, X_2, \ldots, X_j = 1, \ldots, X_k) - p(X_1, X_2, \ldots, X_j = 0, \ldots, X_k)$ - предельный эффект, если Xj – dummy переменная).

Пример вычисления предельного эффекта

```
. sum GRAD ASVABC

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+-------------------------------------------------------
        GRAD |       540    .9425926    .2328351          0          1
      ASVABC |       540    51.36271    9.567646   25.45931   66.07963
```

В рассмотренном примере средний результат A*SVABC* равен 51.36.

```
Logit estimates                                   Number of obs   =        540
                                                  LR chi2(1)      =      43.58
                                                  Prob > chi2     =     0.0000
Log likelihood = -96.886017                       Pseudo R2       =     0.1836

------------------------------------------------------------------------------
        GRAD |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ASVABC |   .1313626    .022428     5.86    0.000     .0874045    .1753206
       _cons |  -3.240218   .9444844    -3.43    0.001    -5.091373   -1.389063
------------------------------------------------------------------------------
```

Пример вычисления предельного эффекта

```
. sum GRAD ASVABC

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------------
        GRAD |       540    .9425926    .2328351          0          1
      ASVABC |       540    51.36271    9.567646   25.45931   66.07963

Logit estimates                                   Number of obs   =        540
                                                  LR chi2(1)      =      43.58
                                                  Prob > chi2     =     0.0000
Log likelihood = -96.886017                       Pseudo R2       =     0.1836

------------------------------------------------------------------------------
        GRAD |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ASVABC |    .1313626    .022428     5.86   0.000      .0874045    .1753206
       _cons |   -3.240218   .9444844    -3.43   0.001     -5.091373   -1.389063
------------------------------------------------------------------------------
```

$Z = -3.240 + 0.131 \times 51.36 = 3.507$

В этой точке *Z* равно 3.507

Пример вычисления предельного эффекта

```
. sum GRAD ASVABC

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        GRAD |       540    .9425926    .2328351          0          1
      ASVABC |       540    51.36271    9.567646   25.45931   66.07963

------------------------------------------------------------------------------
        GRAD |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ASVABC |   .1313626    .022428     5.86   0.000     .0874045    .1753206
       _cons |  -3.240218   .9444844    -3.43   0.001    -5.091373   -1.389063
------------------------------------------------------------------------------
```

$$Z = -3.240 + 0.131 \times 51.36 = 3.507$$

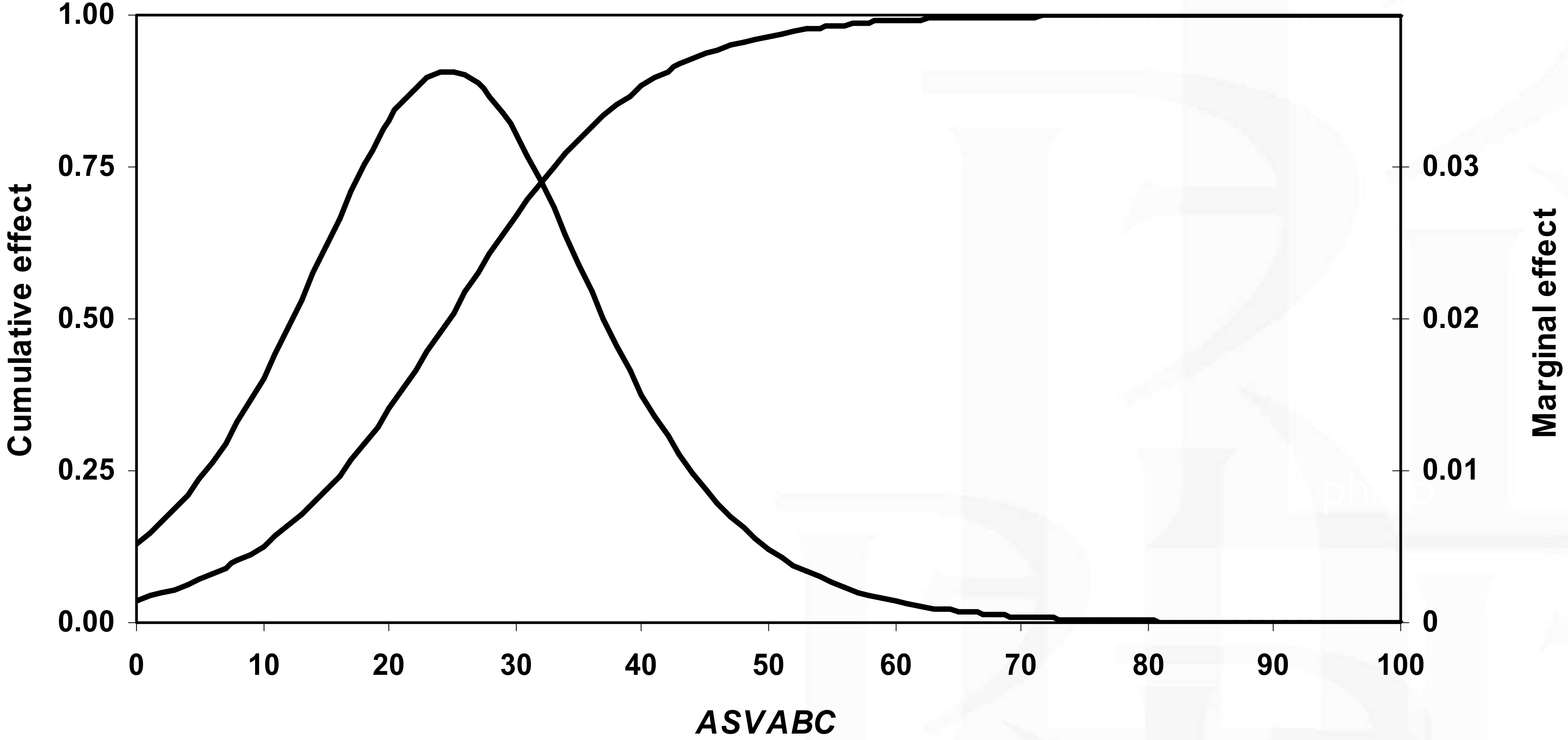
$$e^{-Z} = e^{-3.507} = 0.030$$

$$f(Z) = \frac{dp}{dZ} = \frac{e^{-Z}}{(1+e^{-Z})^2} = \frac{0.030}{(1+0.030)^2} = 0.028$$

$$\frac{\partial p}{\partial X_i} = \frac{dp}{dZ}\frac{\partial Z}{\partial X_i} = f(Z)\beta_i = 0.028 \times 0.131 = 0.004$$

Предельный эффект для имеющего средний результат тестирования равен 0.004. Это означает, что при увеличении результата тестирования *ASVABC* на 1 балл вероятность закончить школу возрастает на 0.4 процента.

Общий случай



Предельный эффект при среднем результате очень мал. Это связано с тем, что вероятность закончить школу при средних результатах и так очень велика.

Интерпретация результатов оценки моделей бинарного выбора

1) **С помощью предельных эффектов (чаще всего в точке средних значений (не очень хорошо для дамми-переменных) или средний эффект по всем наблюдениям)**

2) **С помощью отношения шансов (только для логит-модели)**

3) $P(Y_i = 1) = \frac{1}{1+e^{-z}}$, $P(Y_i = 0) = \frac{e^{-z}}{1+e^{-z}}$,

Отношение шансов (odd ratio) $OR = \frac{P(Y_i = 1)}{P(Y_i = 0)} = e^z = \exp\{\beta_0 + \beta_1 X_1 + \ldots \beta_k X_k\}$,

$Ln(OR) = \beta_0 + \beta_1 X_1 + \ldots \beta_k X_k$

Если X_j изменится на 1 единицу, то OR изменится в $\exp\{\beta_j\}$ раз.

Использование моделей бинарного выбора для прогноза

$\hat{p} = \hat{P}(Y_i = 1) = F(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \ldots \hat{\beta}_k X_k)$ – оценка вероятности, что $Y = 1$.

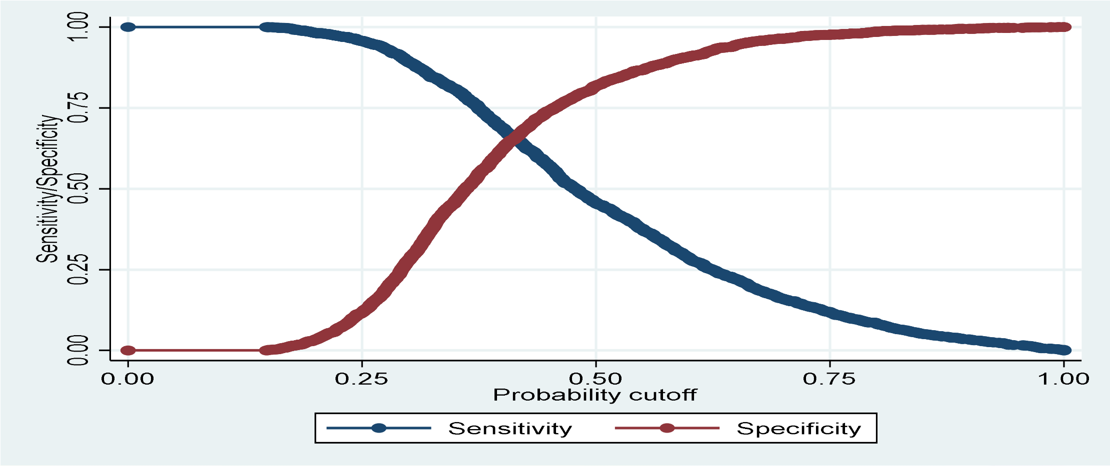
$$\hat{Y}_i = \begin{cases} 1, \text{ если } \hat{p} \geq c \text{ (cutoff)}, \\ 0, \text{ если } \hat{p} < c \end{cases}$$

Обычно $c = 0.5$

Чувствительность (Sensitivity) – доля правильно идентифицированных 1,

Специфичность (Specifity) – доля правильно идентифицированных 0.

Пример:



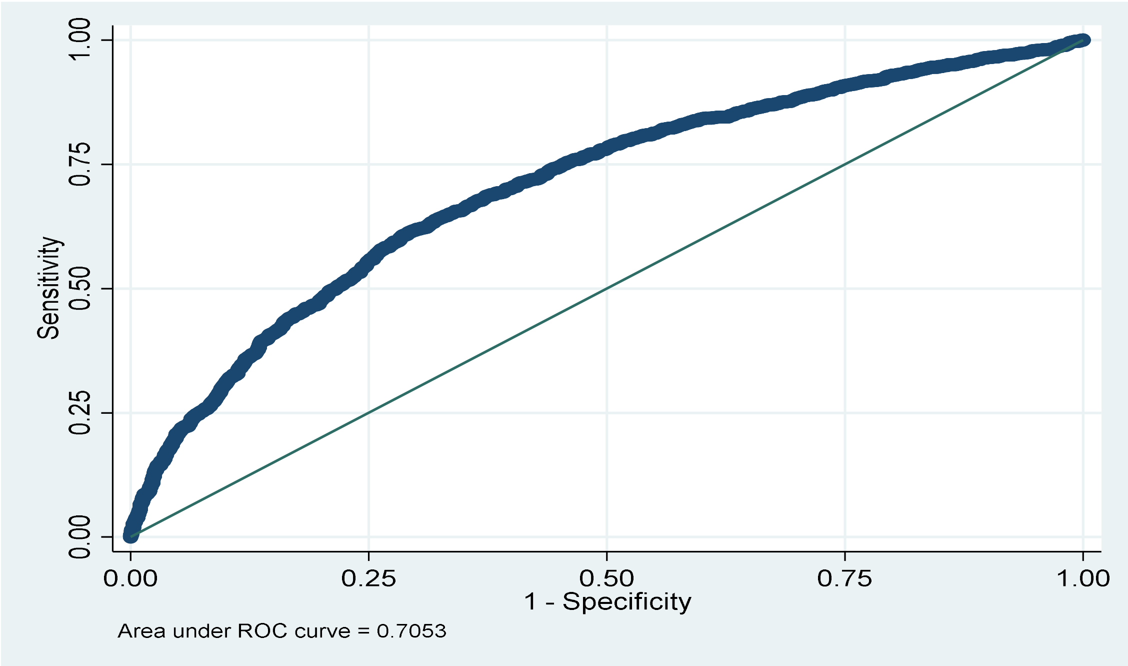
Показатели качества оценки моделей бинарного выбора

Classified	True D	True ~D	Total
+	793	397	1190
-	943	1789	2732
Total	1736	2186	3922

```
Classified + if predicted Pr(D) >= .5
True D defined as male != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   45.68%
Specificity                     Pr( -|~D)   81.84%
Positive predictive value       Pr( D| +)   66.64%
Negative predictive value       Pr(~D| -)   65.48%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   18.16%
False - rate for true D         Pr( -| D)   54.32%
False + rate for classified +   Pr(~D| +)   33.36%
False - rate for classified -   Pr( D| -)   34.52%
--------------------------------------------------
Correctly classified                        65.83%
```

Показатели качества оценки моделей бинарного выбора

ROC – кривая (ROC – received operating characteristics)



AUC = 0.5 – модель не лучше случайного гадания,

AUC < 0.5 – модель хуже случайного гадания

Чувствительность (Sensitivity) – доля правильно идентифицированных 1,

Специфичность (Specifity) – доля правильно идентифицированных 0.

Показатели качества оценки моделей бинарного выбора

R^2_{MF} $-$ R^2 МакФаддена (McFadden) определяется формулой:

R^2_{MF} = 1 - $\frac{\hat{l}}{l_0}$, где $\hat{l}$ - максимум логарифмической функции правдоподобия для рассматриваемой модели, l_0 - максимум логарифмической функции правдоподобия для модели, в которую включена только константа.

Псевдо R^2 (Amemiya, 1981):

$Pseudo\ R^2$ = 1 - $\frac{1}{1+\frac{2}{n}(\hat{l}-l_0)}$

Показатели качества оценки моделей бинарного выбора

Доля неверных прогнозов вычисляется по формуле:

$$wr_1 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.$$

Эта доля неверных прогнозов интересна в сравнении с долей неверных прогнозов самой простой бинарной модели, в которую включена только константа. Для этой просто модели прогнозы строятся совсем просто:

$$\hat{Y}_i = 1, \quad \textit{если} \quad \hat{p} > 0.5, \quad \textit{где} \quad \hat{p} - \textit{доля} \quad \textit{единиц} \quad \textit{в} \quad \textit{выборке},$$
$$\hat{Y}_i = 0, \quad \textit{если} \quad \hat{p} \le 0.5.$$

Доля неверных прогнозов для простой модели вычисляется так:

$$\begin{cases} wr_0 = 1 - \hat{p}, & \textit{если} \quad \hat{p} > 0.5, \\ wr_0 = \hat{p}, & \textit{если} \quad \hat{p} \le 0.5. \end{cases}$$

Отметим, что пороговое значение 0.5 иногда изменяют, чтобы достичь более высоко точности предсказания.

Показатель качества подгонки модели рассчитывается по формуле:

$$R_p^2 = 1 - \frac{wr_1}{wr_0}.$$

Thank you
for your attention!

20, Myasnitskaya str., Moscow, Russia, 101000
Tel.: +7 (495) 628-8829, Fax: +7 (495) 628-7931
www.hse.ru