1 ДЗ-02

1.1 Задача 1.

1.1.1 Logit-модель.

Создадим бинарную переменную `crime86`: 0 - если не был арестован, 1 - если был арестован хоть раз.

```
generate crime86: crime86 = narr86>0
```

Построим логит для `crime86`:

```
logit crime86 pcnv avgsen tottime ptime86 qemp86 inc86 durat black hispan born60
```

```
Logistic regression                             Number of obs     =      2,725
                                                LR chi2(10)       =     253.45
                                                Prob > chi2       =     0.0000
Log likelihood = -1481.4576                     Pseudo R2         =     0.0788

------------------------------------------------------------------------------
     crime86 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        pcnv |  -.9241749   .1244522    -7.43   0.000    -1.168097    -.680253
      avgsen |   .0185028   .0352982     0.52   0.600    -.0506804     .087686
     tottime |    -.01081   .0281724    -0.38   0.701    -.0660269    .0444068
     ptime86 |  -.1230478   .0316273    -3.89   0.000    -.1850361   -.0610595
      qemp86 |   .0664828   .0451872     1.47   0.141    -.0220826    .1550481
       inc86 |  -.0090569   .0012899    -7.02   0.000     -.011585   -.0065287
       durat |   .0196168   .0102139     1.92   0.055    -.0004022    .0396357
       black |   .7763932   .1184542     6.55   0.000     .5442273    1.008559
      hispan |   .4981595   .1099489     4.53   0.000     .2826637    .7136553
      born60 |   .0183557   .0944615     0.19   0.846    -.1667854    .2034967
       _cons |  -.6554766   .1192052    -5.50   0.000    -.8891146   -.4218387
------------------------------------------------------------------------------
```

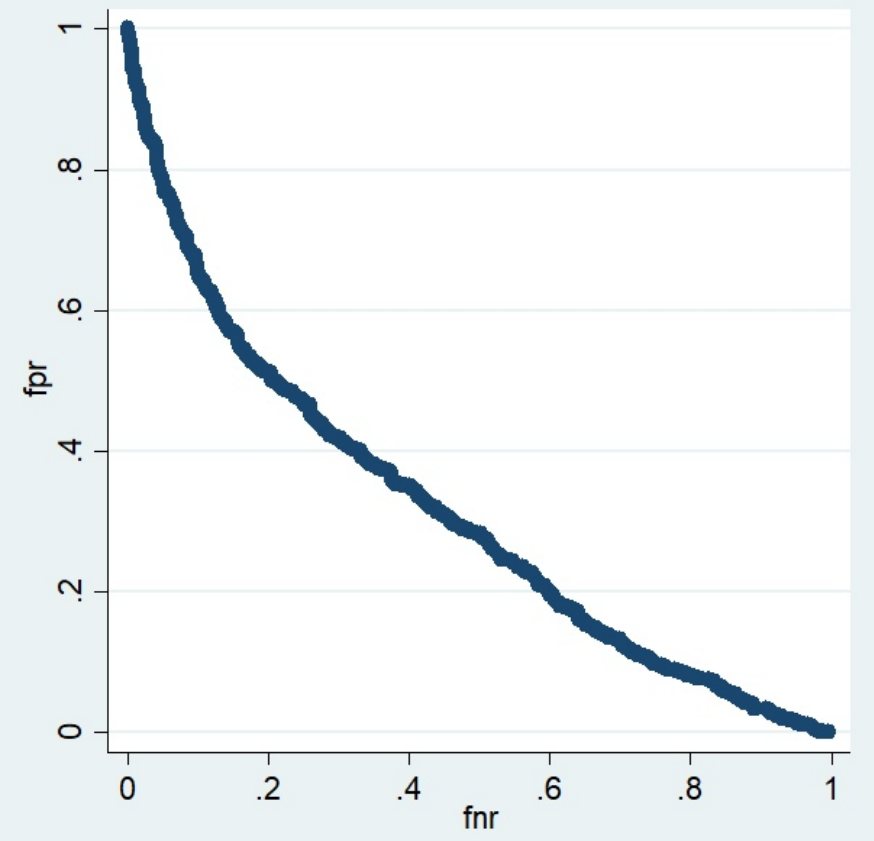
Получим оценку `nocrime` для каждого наблюдения того, что индивидуум не будет ни разу арестован:

```
predict prediction
generate nocrime: nocrime = 1-prediction
```

1.1.2 График долей ошибок I-II рода.

Посчитаем доли ошибок I рода (`fpr`) и II рода (`fnr`) и построим график

```
rocreg crime86 prediction, noboot
generate fpr: fpr = \_fpr\_prediction
generate fnr: fnr = 1-_roc_prediction
twoway (scatter fpr fnr), xsize(5) ysize(5)
```



1.1.3 Poisson count model.

Построим модель Пуассона для `narr86`:

```
poisson narr86 pcnv avgsen tottime ptime86 qemp86 inc86 durat black hispan born60
```

Посчитаем, сколько раз эта модель верно определяет, был ли арестован человек:

```
predict prediction_pois, n
```

```
Poisson regression                              Number of obs     =      2,725
                                                LR chi2(10)       =     387.81
                                                Prob > chi2       =     0.0000
Log likelihood = -2248.0164                     Pseudo R2         =     0.0794

------------------------------------------------------------------------------
      narr86 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        pcnv |  -.4031153   .0849123    -4.75   0.000    -.5695404   -.2366902
      avgsen |  -.0235483   .0199487    -1.18   0.238     -.062647    .0155504
     tottime |   .0243816   .0147567     1.65   0.098    -.0045411    .0533042
     ptime86 |  -.1014503   .0207899    -4.88   0.000    -.1421977    -.060703
      qemp86 |   -.050965   .0309067    -1.65   0.099     -.111541    .0096109
       inc86 |  -.0080571   .0010412    -7.74   0.000    -.0100978   -.0060164
       durat |  -.0080887   .0067295    -1.20   0.229    -.0212782    .0051009
       black |   .6629044   .0738437     8.98   0.000     .5181734    .8076354
      hispan |   .4990338   .0739357     6.75   0.000     .3541225    .6439451
      born60 |  -.0527468    .064075    -0.82   0.410    -.1783316    .0728379
       _cons |  -.5512982   .0777788    -7.09   0.000    -.7037418   -.3988546
------------------------------------------------------------------------------
```

```
count if prediction_pois>0.6 & narr86>=1
244
```

```
count if prediction_pois<= 0.6 & narr86==0
1,723
```

Сравним с логит-моделью:

```
count if prediction>0.4 & crime86==1
226
```

```
count if prediction<=0.4 & crime86==0
1,711
```

Модель Пуассона позволяет получить более точные прогнозы, как для арестованных, так и нет.

В обеих моделях значимые (на 1% уровне значимости) переменные влияют на вероятность ареста с одинаковым знаком.

Зачастую модель регрессии Пуассона плохо соответствует счетным данным, поскольку распределение Пуассона задается единственным параметром (μ). Другим недостатком является то, что модель Пуассона подразумевает равенство дисперсии и математического ожидания, в то время как в счетных данных дисперсия обычно превышает среднее.

Одно из последствий такого однопараметрического моделирования заключается в том, что вероятность нулевых значений, предсказанная по модели Пуассона, значительно ниже, чем их доля в выборке, что называется проблемой избыточных нулевых значений.

Во избежание проблем избыточных нулевых значений и завышенной дисперсии лучше использовать модель бинарного выбора.

1.2 Задача 2.

1.2.1

$x'\hat{\beta}_2 = -0.35474 * 90 - 0.1655 * 9 + 0.2655 * 15.52 + 0 - 0.46766 * 2.9957 + 1.5136 = -0.44697$

$P(at16 = 2)\dfrac{\exp -0.44697}{1 + 0.639564 + 0.33202} = 0.324391$

$x'\hat{\beta}_3 = -0.0451 - 0.29184 * 9 + 0.2189 * 15.52 - 0.78503 * 2.9435 = -1.10256$

$P(at16 = 3)\dfrac{\exp -1.10256}{1 + 0.639564 + 0.33202} = 0.168403$

$P(at16 = 1) = 1 - P(at16 = 2) - P(at16 = 3) = 0.5072064$

1.2.2

$$P(at16|girl) = \frac{0.653572}{1 + 0.653572 + 0.396036} = 0.318877$$

$x'\hat{\beta}_3 = -0.4253$

$\exp(x'\hat{\beta}_3) = 0.653572$

$x'\hat{\beta}_3 = -0.92625$

$\exp x'\hat{\beta}_3 = 0.396036$

$diff = 0.318877 - 0.168403 = 0.150474$

$$\ln \frac{P(at16 = 3|girl)}{P(at16 = 3|boy)} = 0.638447$$

1.2.3

$$\frac{\partial P(at16 = 1)}{\partial loginc} = \left[\frac{1}{1 + \exp(x'\beta_2) + \exp(x'\beta_3)}\right]' =$$
$$= \frac{-1 * (-0.639551 * 0.4676 + 0.33202 * 0.78503)}{(1 + 0.639551 + 0.33202)^2} = 0.144001$$

1.2.4

Независимость от несущественных альтернатив (Independence of irrelevant alternatives) – одна из предпосылок мультиномиальной логистической модели, утверждающая, относительная вероятность выбора между одним из двух вариантов не зависит ни от каких дополнительных альтернатив в наборе, другими словами добавление еще одного элемента к набору вариантов выбора, уменьшит вероятность всех элементов на равную долю.