ДЗ 2 «Бинарная и мультиномиальная логистическая Регрессия»

Выполнил Перетокин Никита (Вариант - Нидерланды)

Уровень значимости(0,95)

Есть ли связь между **участием в выборах** (*vote*) с одной стороны и полом (*gndr*), уровнем доверия политикам (*trstplt-рассмотреть как интервальную*

*переменную*) и удовлетворенностью развитием демократии в стране (*stfdem- рассмотреть как интервальную переменную*). Справедливо ли предположение, что из указанных трех потенциальных предикторов участия в выборах сильнее всего связано с ним доверие политикам?

**Содержательная задача:** Связаны ли факт участия в выборах среди жителей Нидерландов с одной стороны и с другой стороны – полом респондентов, уровнем доверия к политикам и удовлетворенностью развитием демократии.

**Содержательные гипотезы:**

1. Пол респондентов не оказывает влияния на факт участия в выборах
2. Чем выше уровень доверия к политикам, тем выше шанс участия в выборах
3. Чем выше удовлетворенность развитием демократии, тем больше шанс участия в выборах
4. Доверие к политикам оказывает большее влияние на участие в выборах, по сравнению с другими предикторами.

**Регрессия:**

Будет использоваться бинарная логистическая регрессия, т.к. зависимая переменная номинальная и представлена дихотомически, так же необходимо упомянуть, что независимые переменные представлены одной номинальной переменной(пол) и двумя интервальными, поэтому для пола будет создана фиктивная переменная.

**Контрольная группа**

В качестве контрольной группы будут использованы мужчины, которые совершенно не доверяют политикам и совершенно не удовлетворены демократией.

**5. Статистические гипотезы:**

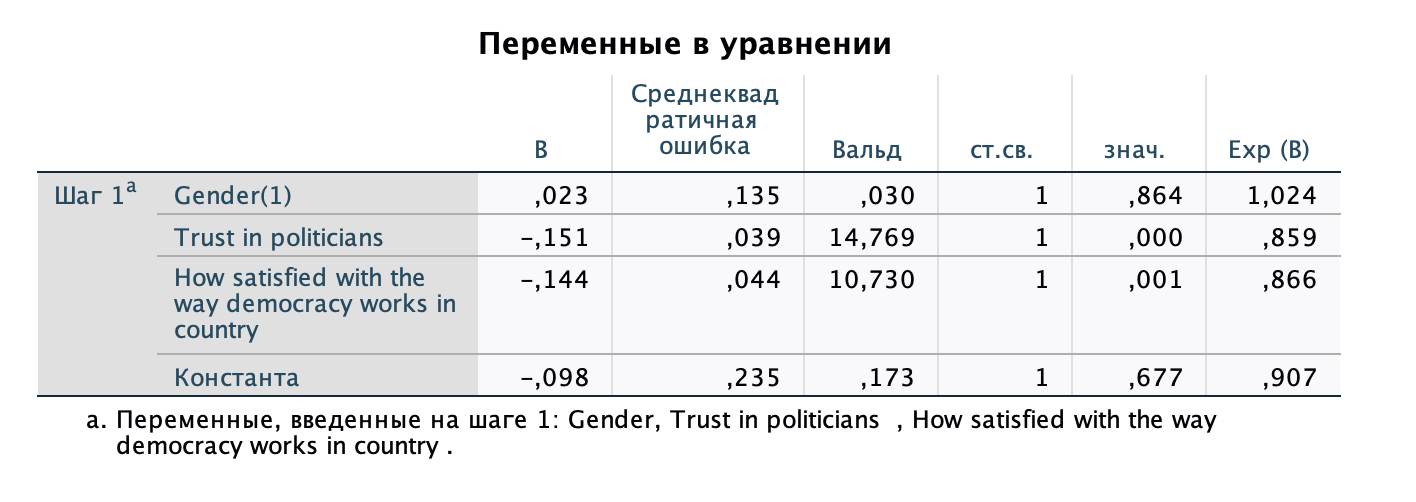
**B0 = 0, expb0 = 1**

**B1 = 0, expb1 = 1**

**B2 = 0, expb2 = 1**

**B3 = 0, expb3 = 1**

**Все расчеты сделаны на доверительном уровне 95%, нет причин менять уровень значимости.**

****

**Проверка статистических гипотез:**

Гипотеза - B0 = 0, expb0 = 1 принимается, это значит, что шансы голосовать и не голосовать на выборах одинаковы для мужчин, которые совершенно не доверяют политикам и совершенно недовольны демократией.

Гипотеза - B1 = 0, expb1 = 1 принимается, это значит, что шансы голосования на выборах не отличаются для женщин и контрольной группы, вне зависимости от других переменных.

B2 = 0, expb2 = 1 не принимается на всех уровнях, это значит, что при увеличении доверия к политикам, шансы не голосовать уменьшаются, вне зависимости от других переменных.

B3 = 0, expb3 = 1 не принимается, это значит, что при увеличении удовлетворенности демократией на одну единицу, шансы не голосовать уменьшаются, вне зависимости от других переменных.

Регрессионное уравнение:

**Интерпретация:**

1 – Для мужчин проживающих в Нидерландах, совершенно не доверяющим политикам и совершенно недовольным демократией, шансы голосовать и не голосовать равны.

1Q1 – Для женщин, проживающих в Нидерландах шансы голосовать, вне зависимости от других переменных не отличаются от контрольной группы.

0,86x1 – при увеличении доверия к политикам на 1 единицу, шансы не проголосовать на выборах уменьшаются в 0,86 раза, вне зависимости от других переменных. (связь обратная)

0,87x2 – при увеличении удовлетворенностью демократией на 1 единицу, шансы не проголосовать уменьшаются в 0.87 раз, вне зависимости от других переменных.(связь обратная)

**Оценка качества модели:**



Здесь проверяется гипотеза H0: внесение независимых переменных не изменяет точность модели.

Значимость в первой таблице говорит о том, что H0 отвергается, значит внесение независимых переменных изменяет точность модели.

Коэффициенты псевдо R2 очень небольшие, что не дает нам оснований считать модель хорошей для предсказаний значений.

Говоря о предиктивных способностях модели, из последней таблицы следует, что для предсказания утвердительных ответов об участии в выборах, модель точна в 100% случаев, и может быть использована для предсказания этих значений.

Однако, для предсказания факта не участия в выборах модель точна в 0% случаев, поэтому не может быть использована для предсказания таких значений.

Общая процентная доля равна 0,86.

**Выводы по задаче и гипотезе:**

Мы обнаружили отсутствие связи между фактом голосования и полом жителей Нидерландов и обнаружили прямую связь между фактом голосования с одной стороны, доверием политикам и удовлетворенностью демократий с другой.

**Гипотеза об отсутствии влияния пола подтвердилась**

Также подтвердилось, что увеличение доверия к политикам и удовлетворенности демократии увеличивает шансы пойти на выборы жителями Нидерландов.

Также подтвердилось, что уровень доверия к политикам, является самым сильным предиктором, хотя его сила не сильно отличается от влияния удовлетворенности демократией.

Задача 2

Есть ли связь между субъективной оценкой здоровья (*health*) с одной стороны и семейным положением (*maritalb*), субъективной оценкой счастья (*happy-рассмотреть как интервальную переменную*). С какой субъективной оценкой здоровья сильнее всего связана оценка степени счастья?

**Задача:**

Выяснить есть ли связь между субъективной оценкой здоровья жителями Нидерландов с одной стороны и с другой – их семейным положением, их субъективной оценкой счастья.

**Гипотезы:**

У легально женатых людей шанс иметь высокую оценку здоровья выше, чем у других групп

Увеличении собственной оценки счастья положительно влияет на оценку собственного здоровья

Оценка степени счастья сильнее связана с очень высокой оценкой здоровья

**Предисловие:**

**Для анализа была выбрана мультиномиальная логистическая регрессия, т.к. зависимая переменная порядковая и представлена не дихотомически.**

**Все расчеты сделаны на доверительном уровне 95%, нет причин менять уровень значимости.**

****

В этой таблице видно, что группы legally separated, in a legally registered civil union, legaly separated и very bad малочисленны, что отразилось на экспонентах, поэтому они были объедены с другими группами: “Bad и Very bad” , “legally married и legally registered civil union” , “legaly separated и legaly divorced”.

**Контрольные группы:**

Для зависимой переменной здоровья, контрольная группа -те люди, которые оценили свое здоровье как плохое или очень плохое, так будет удобно сравнивать переменные.

Для переменной о браке, были выбраны люди, никогда не вступавшие в брак, их было легче противопоставить другим группам.

**Статистические гипотезы:**

**B0 = 0, expb0 = 1**

**B1 = 0, expb1 = 1**

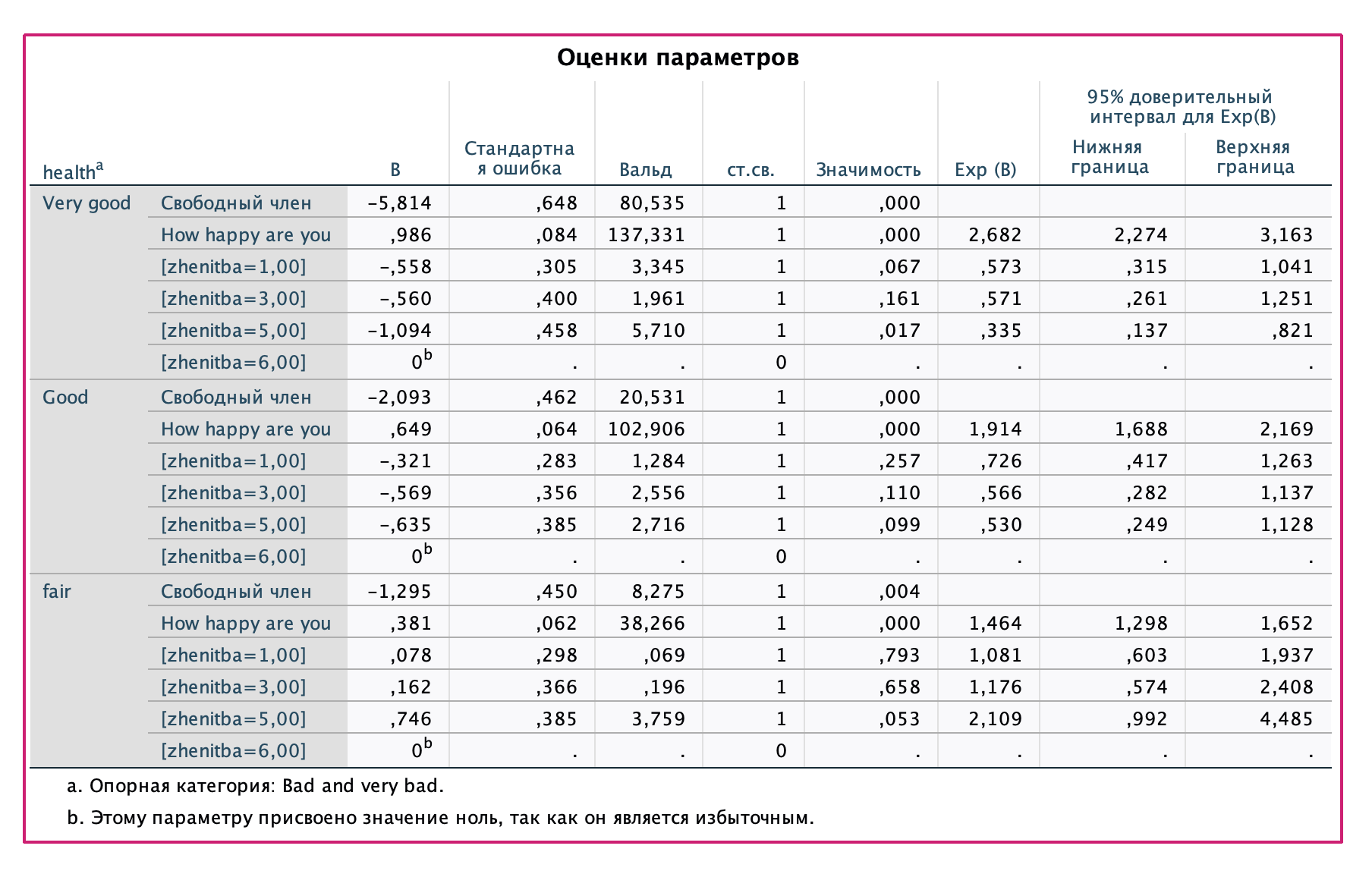
**B2 = 0, expb2 = 1**

**B3 = 0, expb3 = 1**

**B4 = 0, expb4 = 1**

**B5 = 0, expb5 = 1**

Проверка гипотез:



Для зависимой переменной «очень хорошо»:

**B0 = 0, expb0 = 1 – не принимается значимость < 0,05**

**B1 = 0, expb1 = 1 – не принимается значимость < 0,05**

**B2 = 0, expb2 = 1 – принимается на уровне значимость > 0,05**

**B3 = 0, expb3 = 1 – принимается на уровне значимость < 0,05**

**B4 = 0, expb4 = 1 – не принимается на уровне значимость < 0,05**

Для зависимой переменной «хорошо»:

**B0 = 0, expb0 = 1 – не принимается значимость < 0,05**

**B1 = 0, expb1 = 1 – не принимается значимость < 0,05**

**B2 = 0, expb2 = 1 – принимается на уровне значимость > 0,05**

**B3 = 0, expb3 = 1 – принимается на уровне значимость > 0,05**

**B4 = 0, expb4 = 1 – принимается на уровне значимость > 0,05**

Для зависимой переменной «средне»:

**B0 = 0, expb0 = 1 – не принимается значимость < 0,05**

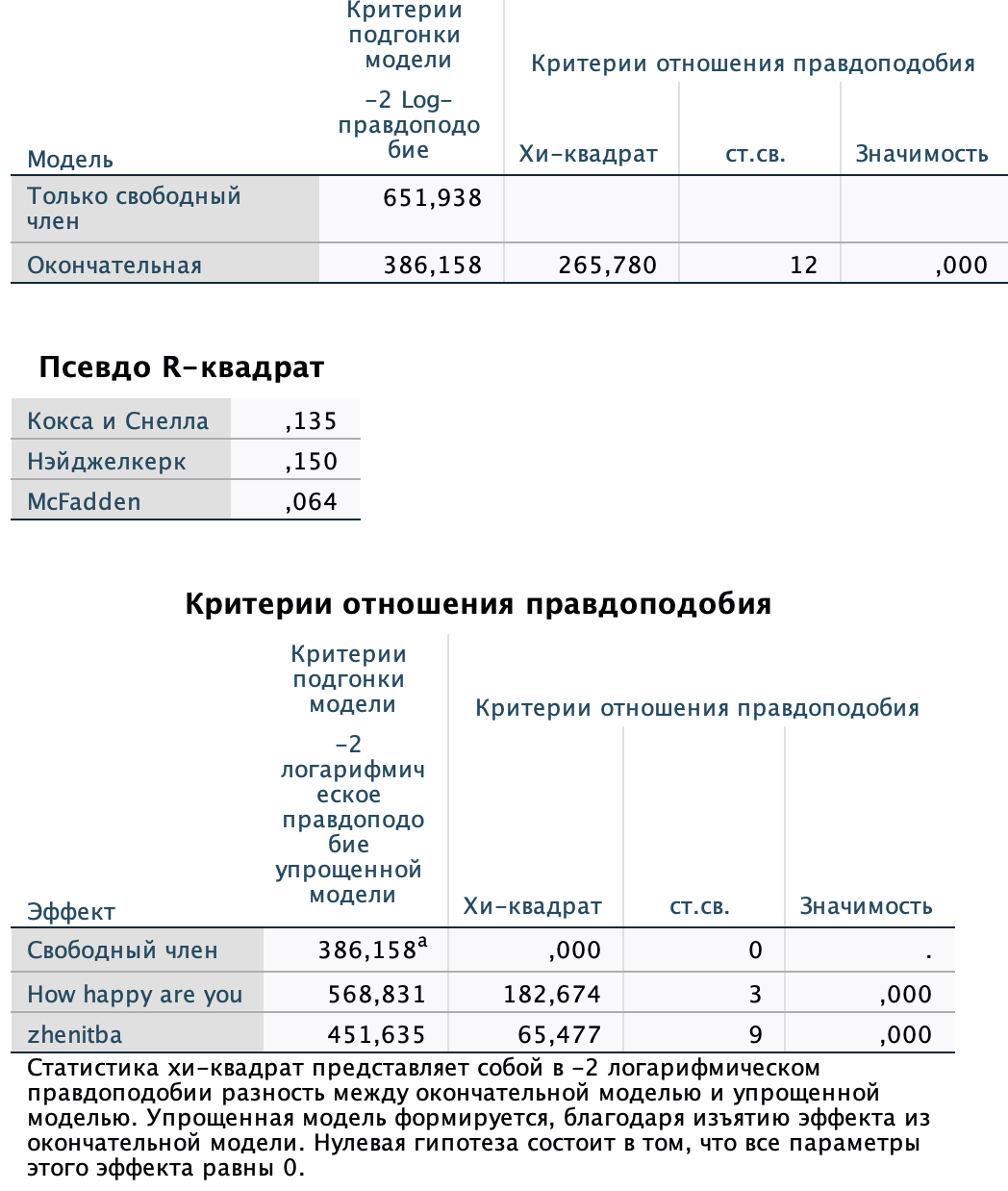
**B1 = 0, expb1 = 1 – не принимается значимость < 0,05**

**B2 = 0, expb2 = 1 – принимается на уровне значимость > 0,05**

**B3 = 0, expb3 = 1 – принимается на уровне значимость > 0,05**

**B4 = 0, expb4 = 1 – принимается на уровне значимость > 0,05**

**Качество модели:**



Здесь проверяется гипотеза H0: внесение независимых переменных не изменяет точность модели.

Значимость в первой таблице говорит о том, что H0 отвергается, значит внесение независимых переменных изменяет точность модели.

Псевдо R2 – все показатели сравнительно не большие, лучше результат дает коэффициент Кокса и Снелла. Эти результаты не дают повода сделать предположение о том, что модель сильна в предсказании значений



Говоря о предиктивных способностях модели, из последней таблицы следует, что модель совершенно не точно в предсказании очень хорошей оценки здоровья, так же малая точность наблюдается для показателей средней, плохо и очень плохой оценок. Однако модель точна для предсказания хороших оценок здоровья.

**Оценка регрессионных уравнений**

Очень хорошо

1. 0,003 – Для людей, никогда не состоящих в браке и оценивающих степень собственного счастья как очень низкое вероятность оценить свое здоровье как очень хорошее меньше, чем оценить свое здоровье как плохое или очень плохое
2. 2,682X1 – При увеличении степени собственного счастья на одну единицу вероятность оценить собственное здоровье как очень хорошее увеличивается в 2,682 раза по сравнению с вероятность оценить свое здоровье как плохое или очень плохое, вне зависимости от других переменных.(Связь прямая)
3. 1Q1 - Для людей, состоящих в каком-либо браке равновероятно оценить свое здоровье как очень хорошее или очень плохо, вне зависимости от других переменных(Связи нет)
4. 1Q2 - Для людей, в разводе и для людей проживающих отдельно от партнера равновероятно оценить свое здоровье как очень хорошее или очень плохо, вне зависимости от других переменных(Связи нет)
5. 0,335Q3 - Для людей, потерявших партнера, вероятность оценить свое здоровье как очень хорошее на 66,5% меньше чем оценить свое здоровье как плохое или очень плохое(Связи обратная)

Хорошо

1. 0,123 – Для людей, никогда не состоявших в браке и оценивающих степень собственного счастья как очень низкое вероятность оценить свое здоровье как хорошее меньше, чем оценить свое здоровье как плохое или очень плохое
2. 1,914X1 – При увеличении степени собственного счастья на одну единицу вероятность оценить собственное здоровье как хорошее увеличивается в 1.914 раза по сравнению с вероятность оценить свое здоровье как плохое или очень плохое, вне зависимости от других переменных.(Связь прямая)
3. 1Q1 - Для людей, состоящих в каком-либо браке равновероятно оценить свое здоровье как хорошее или очень плохо, вне зависимости от других переменных(связи нет)
4. 1Q2 - Для людей, в разводе и для людей проживающих отдельно от партнера равновероятно оценить свое здоровье как хорошее или очень плохо, вне зависимости от других переменных.(Связи нет)
5. 1Q3 - Для людей, потерявших партнера, равновероятно оценить свое здоровье как хорошее или очень плохо, вне зависимости от других переменных.(Связи нет)

Средне

1. 0,273 – Для людей, никогда не состоящих в браке и оценивающих степень собственного счастья как очень низкое вероятность оценить свое здоровье как очень хорошее на 0,727 меньше чем оценить свое здоровье как плохое или очень плохое
2. 1,298X1 – При увеличении степени собственного счастья на одну единицу вероятность оценить собственное здоровье как среднее увеличивается в 1,298 раза по сравнению с вероятность оценить свое здоровье как плохое или очень плохое, вне зависимости от других переменных.(связь прямая)
3. 1Q1 - Для людей, состоящих в каком-либо браке равновероятно оценить свое здоровье как очень хорошее или очень плохо, вне зависимости от других переменных(связи нет)
4. 1Q2 - Для людей, в разводе и для людей проживающих отдельно от партнера равновероятно оценить свое здоровье как среднее хорошее или очень плохо, вне зависимости от других переменных(связи нет)
5. 1Q3 - Для людей, потерявших партнера, равновероятно оценить свое здоровье как среднее или очень плохо, вне зависимости от других переменных(связи нет)

**Интерпретация:**

Мы выяснили наличие связи между субъективной оценкой здоровья жителями Нидерландов с одной стороны и с другой – их семейным положением, их субъективной оценкой счастья. Стоит заметить, что оценка собственного счастья оказывает большее влияние на очень хорошую оценку здоровья.

**Гипотезы:**

**У легально женатых людей шанс иметь высокую оценку здоровья выше, чем у других групп**

Не подтвердилась, шансы иметь высокую, среднюю и очень высокую оценку здоровья людьми состоящих в браках не отличалось от шансов иметь оценку плохую или очень плохую

**Увеличении собственной оценки счастья положительно влияет на оценку собственного здоровья**

Подтвердилась, В группах с очень хорошей и хорошей оценками здоровья, вероятность повышалась при увеличении собственной оценки счастья

**Оценка степени счастья сильнее связана с очень высокой оценкой здоровья**

Подтвердилась, при увеличении степени собственного счастья на одну единицу вероятность оценить собственное здоровье как очень хорошее увеличивается в 2,682 раза по сравнению с контрольной группы.

**Приложение:**

**Задание 1**

Отбор наблюдений

USE ALL.

COMPUTE filter\_$=(cntry = 'NL').

VARIABLE LABELS filter\_$ "cntry = 'NL' (FILTER)".

VALUE LABELS filter\_$ 0 'Not Selected' 1 'Selected'.

FORMATS filter\_$ (f1.0).

FILTER BY filter\_$.

EXECUTE.

Построение регрессии

LOGISTIC REGRESSION VARIABLES vote

/METHOD=ENTER gndr trstplt stfdem

/CONTRAST (gndr)=Indicator(1)

/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

Задание 2

Проверка наполненности групп

NOMREG health (BASE=LAST ORDER=ASCENDING) BY maritalb WITH happy

/CRITERIA CIN(95) DELTA(0) MXITER(100) MXSTEP(5) CHKSEP(20) LCONVERGE(0) PCONVERGE(0.000001)

SINGULAR(0.00000001)

/MODEL

/STEPWISE=PIN(.05) POUT(0.1) MINEFFECT(0) RULE(SINGLE) ENTRYMETHOD(LR) REMOVALMETHOD(LR)

/INTERCEPT=INCLUDE

/PRINT=CLASSTABLE PARAMETER SUMMARY LRT CPS STEP MFI.

Перекодировка переменных

RECODE maritalb (5=5) (6=6) (1 thru 2=1) (3 thru 4=3) INTO marii.

VARIABLE LABELS marii 'zhenitba'.

EXECUTE.

RECODE health (1=1) (2=2) (3=3) (4 thru 5=4) INTO heath1.

VARIABLE LABELS heath1 'health'.

EXECUTE.

Построение регрессии

NOMREG heath1 (BASE=LAST ORDER=ASCENDING) BY marii WITH happy

/CRITERIA CIN(95) DELTA(0) MXITER(100) MXSTEP(5) CHKSEP(20) LCONVERGE(0) PCONVERGE(0.000001)

SINGULAR(0.00000001)

/MODEL

/STEPWISE=PIN(.05) POUT(0.1) MINEFFECT(0) RULE(SINGLE) ENTRYMETHOD(LR) REMOVALMETHOD(LR)

/INTERCEPT=INCLUDE

/PRINT=CLASSTABLE PARAMETER SUMMARY LRT CPS STEP MFI.