**Федеральное государственное автономное образовательное учреждение высшего образования**

**«Национальный исследовательский университет**

**«Высшая школа экономики»»**

**Департамент социологии**

**ДОМАШНЕЕ ЗАДАНИЕ №3**

по дисциплине

«Анализ данных в социологии» на тему:

Регрессия с фиктивными переменными

Вариант: «Нидерланды»

Выполнил:

Студент группы БСЦ-194

*Перетокин Никита Александрович*

Преподаватель:

Доцент

*Зангиева Ирина Казбековна*

Москва 2021

**Задание 1.1.1**

**Содержательная задача:**

Выяснить как половая принадлежность и оценка собственного счастья влияет на оценку степени счастья среди жителей Нидерландов независимо друг от друга.

**Содержательная гипотеза:**

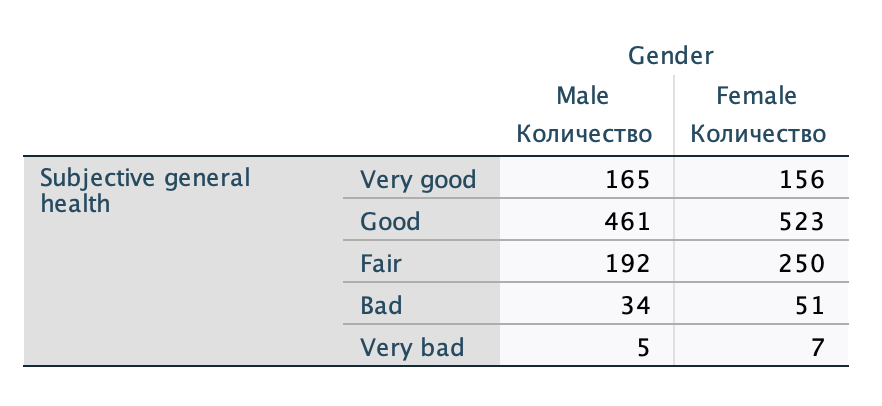
1. Можно предположить, что чем лучше респондент оценил состояние своего здоровья, тем выше его оценка собственного счастья, т.к. плохое здоровье мешает наслаждать жизнью.
2. Пол не оказывает влияние на оценку собственного счастья жителей Нидерландов

**Вид регрессии:**

Для этого задания будет использоваться линейная регрессия с двумя фиктивными переменными, т.к. Переменная пол – номинальная, а оценка собственного здоровья – порядковая. Поэтому будет использоваться независимое кодирование

В качестве контрольной группы были выбраны мужчины, оценивающие свое здоровье как среднее. Такой выбор обоснован тем, что нам интересно сравнить влияния негативной и позитивной оценки в равной степени. Удобнее если, переменная, которая будет использована для сравнения, будет лежать по середине от двух крайностей.

Контрольная группа представлена 165 людьми:



**Статистические Гипотезы:**

H0: R2=0, Модель не объясняет связь переменных; H1: R2 ≠ 0

H0: B0 =0.; H1: B0 ≠0

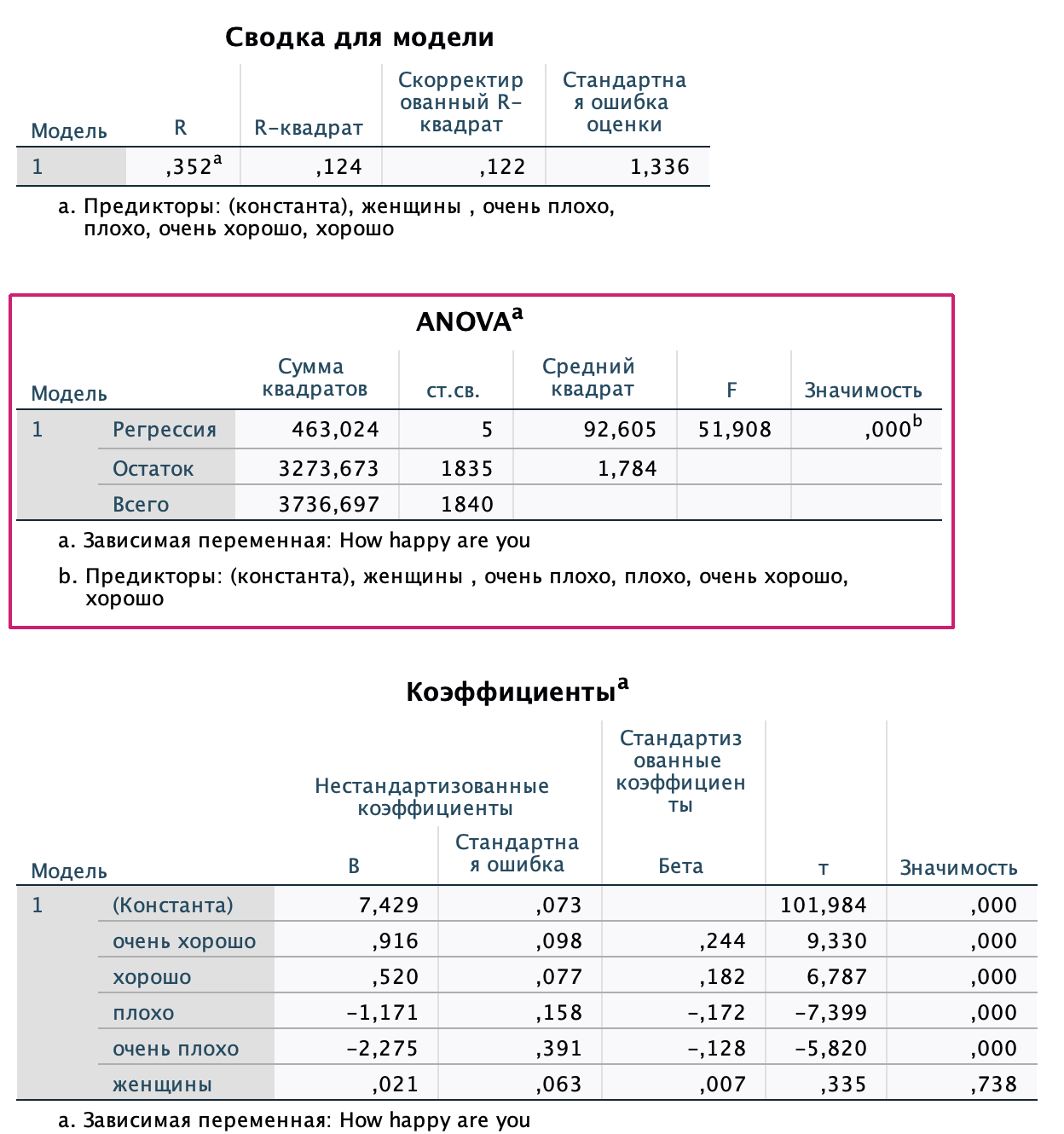
H0: B1 =0. Оценка здоровья как очень хорошее не оказывает влияния на оценку счастья; H1: B1≠0

H0: B2 =0. Оценка здоровья как хорошее не оказывает влияния на оценку счастья; H1: B2 ≠0

H0: B3 =0. Оценка здоровья как плохое не оказывает влияния на оценку счастья; H1: B3 ≠0

H0: B4 =0. Оценка здоровья как очень плохое не оказывает влияния на оценку счастья; H1: B4 ≠0

H0: B5 =0. Пол респондента не оказывает влияния на оценку счастья; H1: B5 ≠0



Для проверки был выбран общепринятый доверительны уровень 95%, потому что выборка сравнительно большая

**Проверка статистических гипотез:**

1. Значимости коэффициента детерминация стремится к нулю, следовательно гипотеза об отсутствии связи R2=0 не принимается, в пользу гипотезы H1: R2 ≠ 0. Однако, коэффициент R2  равен 0,124, это означает, что модель описывает лишь 12,4% наблюдений, то есть модель не может быть использована для прогноза.
2. Гипотеза H0: B5 =0 о значимости пола принимается, это значит, что половая принадлежность не влияет на собственную оценку счастья.
3. Все остальные гипотезы не принимаются, это означает, что оценка здоровья влияет на оценку счастья респондента.

**Регрессионное уравнение:**

7,429 + 0,916\*(q1) + 0,520\*(q2) -1,171\*(q3)-2,275(q4)

**Интерпретация:**

1. **B0 = 7,429.** Другими словами, средняя оценка счастья, среди мужчин, проживающих в Нидерландах, оценивающих свое здоровье как среднее равна 7,429.
2. **B1 = 0,916.** Средняя оценка счастья среди жителей Нидерландов, которые оценивают свое здоровье как очень хорошее больше на 0,916 по сравнению с контрольной группой и равна 8,345.
3. **B2 = 0, 52.** Средняя оценка счастья среди жителей Нидерландов, которые оценивают свое здоровье как хорошее больше на 0,52 по сравнению с контрольной группой и равна 6,258.
4. **B3 = -1,171.** Средняя оценка счастья среди жителей Нидерландов, которые оценивают свое здоровье как плохое меньше на 1,171 по сравнению с контрольной группой и равна 6,258.
5. **B4 = -2,275.** Средняя оценка счастья среди жителей Нидерландов, которые оценивают свое здоровье как очень плохое меньше на 2,275 по сравнению с контрольной группой и равна 4,670.
6. **B5 = 0.** Средняя оценка счастья среди жителей Нидерландов женского пола не отличается от контрольной группой и равна 7,429.

**Содержательные выводы:**

Была обнаружена прямая линейная связь, другими словами, чем лучше респондент оценивает свое здоровье, тем, в среднем выше его оценка счастья. Также обнаружено, что пол не влияет на оценку собственного счастья

**Гипотезы:**

1. Гипотеза о влиянии оценки собственного здоровья среди жителей Нидерландов на собственную оценку счастья подтвердилась.
2. Гипотеза об отсутствии влияния пола на оценку счастья среди жителей Нидерландов подтвердилась.

**Задача 1.1.2**

**Содержательная задача:**

Выявить как связана оценка счастья жителями Нидерландов и сочетание их пола и оценки собственного здоровья.

**Содержательные гипотезы:**

1. Для обеих половых групп имеет место прямая связь между оценкой счастья

И собственной оценкой здоровья.

1. Для обоих половых групп оценка собственного здоровья оказывает одинаковое влияние на собственной оценкой счастья.

**Метод:**

Будет использоваться Регрессионная модель с фиктивными переменными, Кодировка будет совместной, Т.К. нам необходимо выявить взаимодействие собственной оценки здоровья и половой принадлежности.

**Контрольная группа:**

Контрольная группа остается прежней, для удобного сравнения двух методов, а именно – мужчины, оценивающие свое здоровье как среднее.

**Статистические гипотезы:**

H0: R2=0; H1: R2 ≠ 0

H0: B0 =0.; H1: B0 ≠0

H0: B1 =0; H1: B1 ≠0

H0: B2 =0; H1: B2 ≠0

H0: B3 =0; H1: B3 ≠0

H0: B4 =0; H1: B4 ≠0

H0: B5 =0; H1: B5 ≠0

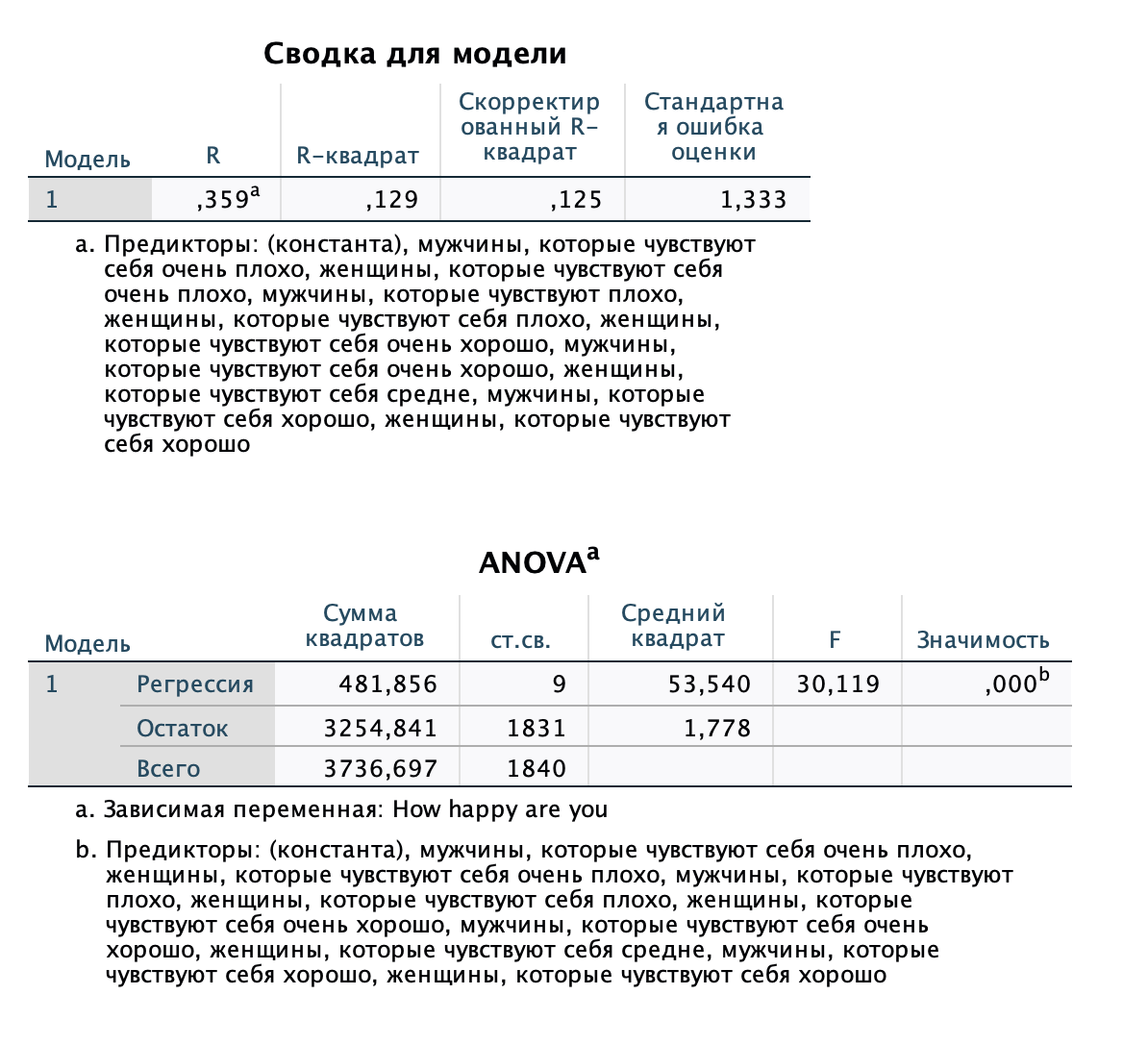
H0: B6 =0; H1: B6 ≠0

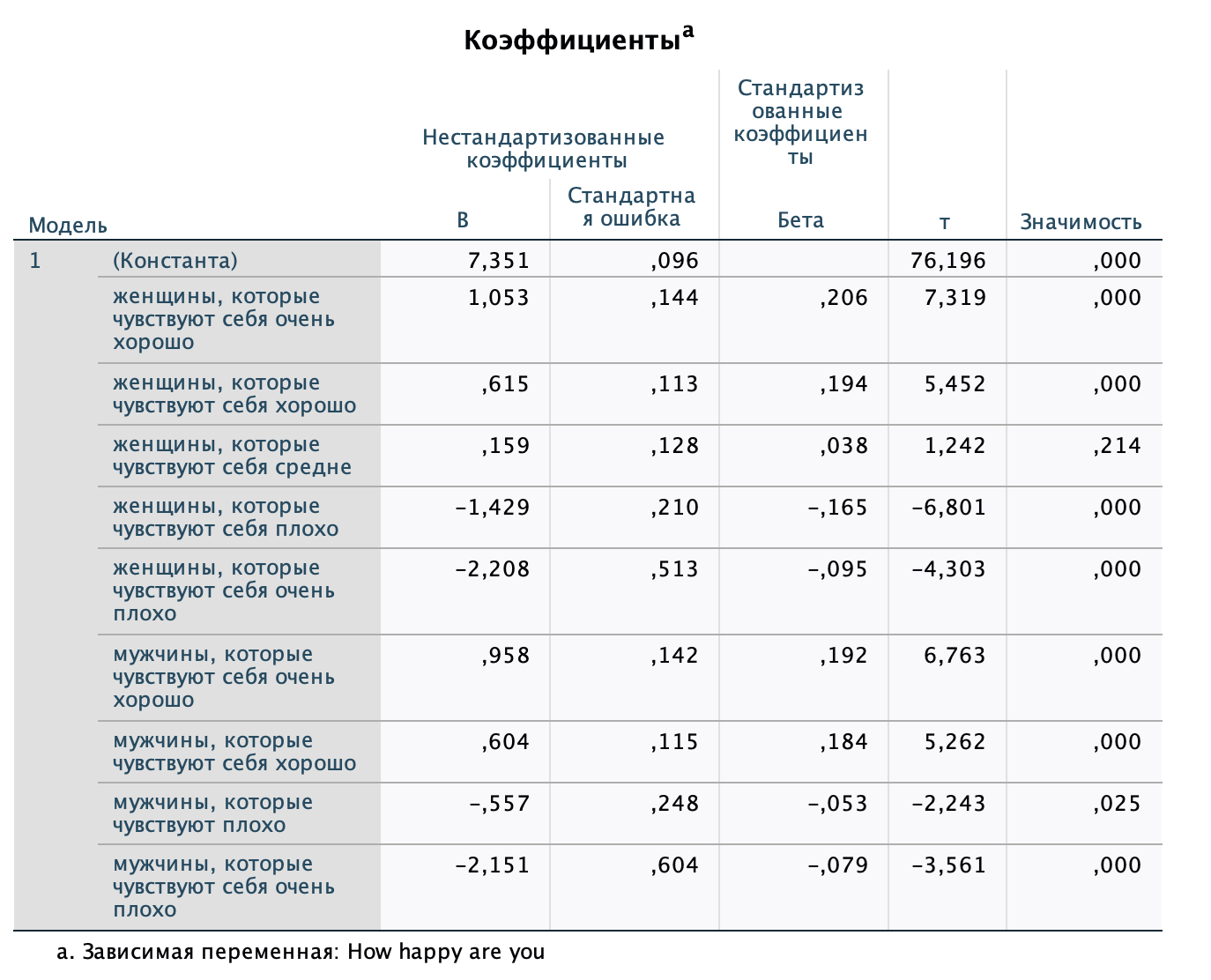
H0: B7 =0; H1: B7 ≠0

H0: B7 =0; H1: B7 ≠0

H0: B8 =0; H1: B8 ≠0

H0: B9 =0; H1: B9 ≠0

****

****

**Доверительный интервал был выбран 95%:**

1. Значимости коэффициента детерминация стремится к нулю, следовательно гипотеза об отсутствии связи R2=0 не принимается, в пользу гипотезы H1: R2 ≠ 0. Однако, коэффициент R2  равен 0,129, это означает, что модель описывает лишь 12,9% наблюдений, то есть модель не может быть использована для прогноза.
2. Все статистические гипотезы не принимаются, кроме гипотезы

H0: B3 =0 об отсутствии влияние сочетания женского пола и средней оценки собственного здоровья на оценку счастья респондента.

**Уравнение регрессии:**

Y = 7,351+ 1,053\*(q1) + 0,615\*(q2) + 0,159 \* (q3) – 1,429\*(q4) -2,209\*(q5)

+ 0,958\*(q6) + 0,604\*(q7) – 0,557\*(q8) – 2,151\*(q9)

Т.К. гипотеза H0: B3 =0 принимается логично переписать уравнение как:

Y = 7,351+ 1,053\*(q1) + 0,615\*(q2) – 1,429\*(q4) -2,209\*(q5)

+ 0,958\*(q6) + 0,604\*(q7) – 0,557\*(q8) – 2,151\*(q9)

**Интерпретация коэффициентов:**

1. **B0 = 7,351.** Другими словами, средняя оценка счастья, среди мужчин, проживающих в Нидерландах, оценивающих свое здоровье как среднее равна 7,351.
2. **B1 = 1,053.** Средняя оценка счастья среди жителей Нидерландов, женского пола, которые оценивают свое здоровье как очень хорошее больше на 1,053 по сравнению с контрольной группой и равна 8,404.
3. **B2 = 0,615.** Средняя оценка счастья среди жителей Нидерландов, женского поло, которые оценивают свое здоровье как хорошее больше на 0,615 по сравнению с контрольной группой и равна 7,966.
4. **B3 = 0.** Средняя оценка счастья среди жителей Нидерландов, женского поло, которые оценивают свое здоровье как хорошее не отличается от контрольной группой и равна 7,351.
5. **B4 = -1,429.** Средняя оценка счастья среди жителей Нидерландов, женского пола, которые оценивают свое здоровье как плохое меньше на 1,429 по сравнению с контрольной группой и равна 5,992.
6. **B5 = -2,209.** Средняя оценка счастья среди жителей Нидерландов женского пола, которые оценивают свое здоровье как очень плохое ниже на 2,209 по сравнению с контрольной группой и равна 5,142.
7. **B6 = 0,958** Средняя оценка счастья среди жителей Нидерландов, мужского пола, которые оценивают свое здоровье как очень хорошее больше на 0,958 по сравнению с контрольной группой и равна 8,309.
8. **B7 = 0, 604.** Средняя оценка счастья среди жителей Нидерландов, мужского пола, которые оценивают свое здоровье как хорошее больше на 0,604 по сравнению с контрольной группой и равна 7,955.
9. **B4 = -0,557.** Средняя оценка счастья среди жителей Нидерландов, мужского пола, которые оценивают свое здоровье как плохое меньше на 0,557 по сравнению с контрольной группой и равна 6,794.
10. **B6 = -2,151** Средняя оценка счастья среди жителей Нидерландов, мужского пола, которые оценивают свое здоровье как очень плохое меньше на 2,151 по сравнению с контрольной группой и равна 5,2.

**Содержательные выводы:**

Было выявлено линейное влияние сочетания половой принадлежности и оценки собственного здоровья на оценку счастья среди жителей Нидерландов.

Можно утверждать, что влияние половой принадлежности менее значимо, этот вывод можно сделать исходя из того, что группа Q3(Женщины, оценивающие свое здоровье как среднее) был признан не значимым, оценка здоровья одинакова, но изменение пола не принесла значимых изменений в оценки счастья.

Однако, нельзя утверждать, что сила влияние оценки собственного здоровья одинаково для мужчин и женщин, если сравнить оценки счастья среди мужчин и женщин, можно прийти к выводу, что оценка собственного здоровья оказывает больше влияние на оценку счастья среди женщин, это проявляется во все группах оценки здоровья.

**Гипотезы:**

Гипотеза 1 подтверждается, имеет место прямая связь между собственной оценкой счастья и совокупностью половой принадлежности с оценкой собственного здоровья дли жителей Нидерландов.

Гипотеза 2 не подтверждается, было замечено, что оценка собственного здоровья среди жителей Нидерландов оказывает большее влияние на собственную оценку счастья среди женщин.

**Задача 1.2.**

Более целесообразно выбрать модель с совместной кодировкой, она дает нам дополнительные выводы, а именно указывает на отличия влияния оценки здоровья среди женщин и мужчин.

Таким образом выводы о наличии линейной связи, сделанные в первой модели подтверждаются второй, но также к ним добавляются дополнительные результаты.

**Задача 2**

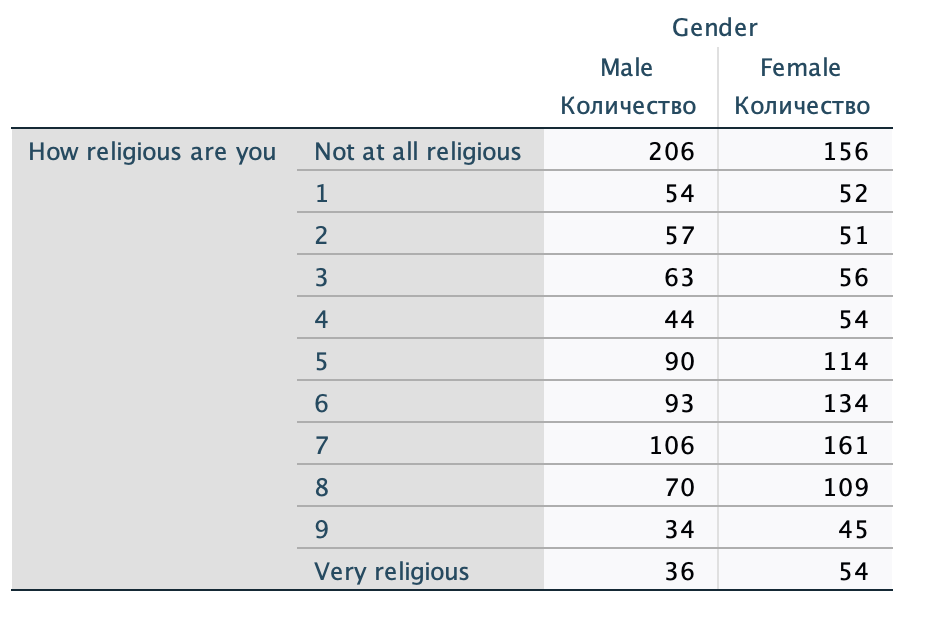
**Содержательная задача:**

Выявить связь пола, степени религиозности с оценкой степени счастья.

Выяснить одинакова ли сила связи между степенью религиозности и оценкой степени счастья среди мужчин и женщин, проживающих в Нидерландах.

**Содержательные гипотезы:**

1. Среди жителей Нидерландов наблюдается линейная положительная связь между степенью религиозности и личной оценкой счастья
2. Сила связи между степенью религиозности и оценкой степени счастья не зависит от пола.

****

**Метод:**

Будет использована регрессионная модель смешного типа, будут использованы как и фиктивные, так и интервальные переменные, Т.К. переменная «религия» является пвсевдоинтервальной, а переменная «пол» номинальной.

**Контрольная группа:**

В качестве контрольной группы были выбраны мужчины(чтобы не создавать новые переменные). Группа состоит из 853 человека

**Статистические гипотезы:**

H0: R2=0; H1: R2 ≠ 0

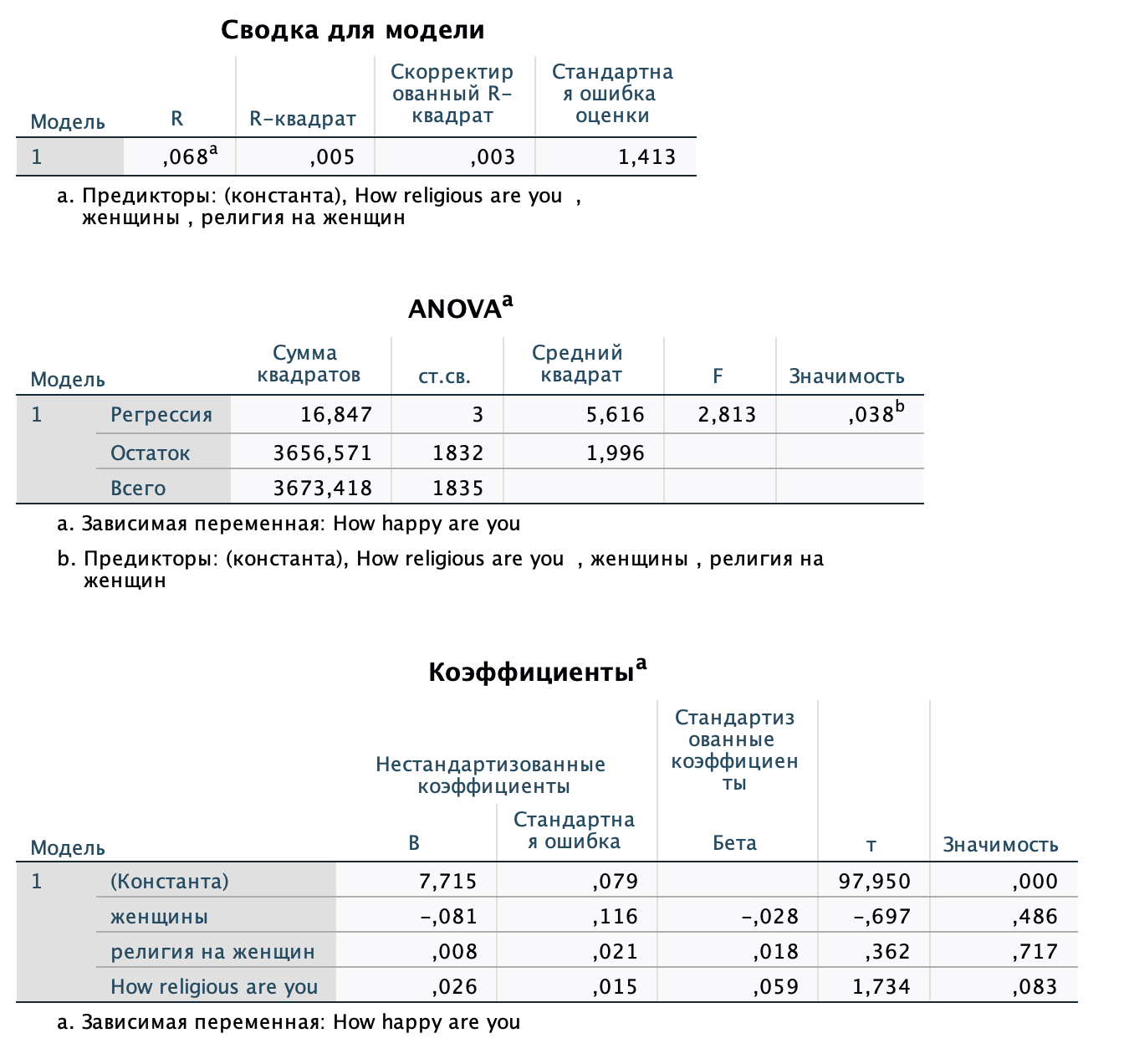
H0: B0 =0.; H1: B0 ≠0

H0: B1 =0; H1: B1 ≠0

H0: B2 =0; H1: B2 ≠0

H0: B3 =0; H1: B3 ≠0

**В качестве Доверительного интервала был выбран уровень 90%, большая выборка позволяет его понизить**

****

1. Значимости коэффициента детерминация стремится к нулю, следовательно гипотеза об отсутствии связи R2=0 не принимается, в пользу гипотезы H1: R2 ≠ 0. Однако, коэффициент R2  равен 0,005, это означает, что модель описывает лишь 0,5% наблюдений, то есть модель не может быть использована для прогноза.
2. Гипотезы H0: B1 =0; и B2 =0; H0: B2 =0 принимаются.
3. Гипотеза H0: B3 =0 не принимается в пользу гипотезы H1: B3 ≠0

**Регрессионное уравнение:**

**Y = 7,715 + 0,026\*(q3)**

**Интерпретация коэффициентов:**

1. B0 = 7,715. В среднем мужчины, проживающие в Нидерландах которые совершенно не религиозны оценивают свое счастье на 7,715. Что является довольно высоким показателем
2. B1 = 0. Это означает, что средняя оценка счастья среди женщин не отличается от средней оценки счастья мужчин.
3. B2 = 0, это означает, что при увеличении оценки религиозности среди женщин на 1, средний показатель счастья не меняется. Таким образам отличия в силе связи для двух полов не обнаружена.
4. B3 = 0,026. При увеличении религиозности на 1, средняя оценка счастья для обоих полов, в среднем увеличивается на 0,026. Обнаружена прямая положительная связь7

**Содержательные выводы:**

Была выявлена прямая положительная связь между степенью религиозности и собственной оценки счастья среди жителей Нидерландов, вне зависимости от половой принадлежности.

Также было обнаружено отсутствие связи между полом и собственной оценкой счастья.

К тому же было выяснено, что сила связи между религиозностью и собственной оценкой счастья среди жителей Нидерландов, не зависит от половой принадлежности.

**Таким образом:**

1. Гипотеза 1 о тот, что Среди жителей Нидерландов наблюдается линейная положительная связь между степенью религиозности и личной оценкой счастья была подтверждена.
2. Гипотеза о том, что Сила связи между степенью религиозности и оценкой степени счастья не зависит от пола подтвердилась.

**Приложение:**

**Отбор жителей Нидерландов:**

**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.**

**1.1.1**

Строим таблицу «Оценка здоровья» - «Пол»

CTABLES

/VLABELS VARIABLES=health gndr DISPLAY=LABEL

/TABLE health BY gndr [COUNT F40.0]

/CATEGORIES VARIABLES=health gndr ORDER=A KEY=VALUE EMPTY=INCLUDE MISSING=EXCLUDE

/CRITERIA CILEVEL=95.

**Создаем фиктивную перемену для женщин**

RECODE gndr (2=1) (SYSMIS=0) (1 =0)INTO fe.

VARIABLE LABELS fe 'женщины '.

EXECUTE.

Создаем фиктивные переменные для оценки здоровья

RECODE health (1=1) (SYSMIS=0) (2 thru 5=0) INTO h1.

VARIABLE LABELS h1 'очень хорошо'.

EXECUTE.

RECODE health (2=1) (SYSMIS=0) (3 thru 5=0) (1 = 0) INTO h2.

VARIABLE LABELS h2 'хорошо'.

EXECUTE.

RECODE health (4=1) (SYSMIS=0) (4 thru 5=0) (1 thru 3=0)INTO h3.

VARIABLE LABELS h3 'плохо'.

EXECUTE.

RECODE health (5=1) (SYSMIS=0) (1 thru 4=0) INTO h4.

VARIABLE LABELS h4 'очень плохо'.

EXECUTE.

Строим регрессию

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.

REGRESSION

/MISSING LISTWISE

/STATISTICS COEFF OUTS R ANOVA

/CRITERIA=PIN(.05) POUT(.10)

/NOORIGIN

/DEPENDENT happy

/METHOD=ENTER h1 h2 h3 h4 fe.

1.1.2

Создаем фиктивные переменные для каждого сочетания, кроме кг

if health=1 and gndr=2 q1=1.

VARIABLE LABELS q1 'женщины, которые чувствуют себя очень хорошо'.

EXECUTE.

if health=2 and gndr=2 q2=1.

VARIABLE LABELS q2 'женщины, которые чувствуют себя хорошо'.

EXECUTE.

if health=3 and gndr=2 q3=1.

VARIABLE LABELS q3 'женщины, которые чувствуют себя средне'.

EXECUTE.

if health=4 and gndr=2 q4=1.

VARIABLE LABELS q4 'женщины, которые чувствуют себя плохо'.

EXECUTE.

if health=5 and gndr=2 q5=1.

VARIABLE LABELS q5 'женщины, которые чувствуют себя очень плохо'.

EXECUTE.

if health=1 and gndr=1 q6=1.

VARIABLE LABELS q6 'мужчины, которые чувствуют себя очень хорошо'.

EXECUTE.

if health=2 and gndr=1 q7=1.

VARIABLE LABELS q7 'мужчины, которые чувствуют себя хорошо'.

execute.

if health=4 and gndr=1 q8=1.

VARIABLE LABELS q8 'мужчины, которые чувствуют плохо'.

EXECUTE.

if health=5 and gndr=1 q9=1.

VARIABLE LABELS q9 'мужчины, которые чувствуют себя очень плохо'.

EXECUTE.

Заполняем пропуски

DO IF ((health <= 5) & (gndr <= 2)).

RECODE q1 q2 q3 q4 q5 q6 q7 q8 q9 (SYSMIS=0).

END IF.

EXECUTE.

Строим регрессию

REGRESSION

/MISSING LISTWISE

/STATISTICS COEFF OUTS R ANOVA

/CRITERIA=PIN(.05) POUT(.10)

/NOORIGIN

/DEPENDENT happy

/METHOD=ENTER q1 q2 q3 q4 q5 q6 q7 q8 q9.

**2**

**Строим таблицу «религиозность» - «пол»**

CTABLES

/VLABELS VARIABLES=rlgdgr gndr DISPLAY=LABEL

/TABLE rlgdgr BY gndr [COUNT F40.0]

/CATEGORIES VARIABLES=rlgdgr gndr ORDER=A KEY=VALUE EMPTY=INCLUDE MISSING=EXCLUDE

/CRITERIA CILEVEL=95.

**Создаем совместную переменную для женщин и религиозности**

COMPUTE R1=rlgdgr \* fe.

VARIABLE LABELS R1 'религия на женщин'.

EXECUTE.

**Строим регрессию**

REGRESSION

/MISSING LISTWISE

/STATISTICS COEFF OUTS R ANOVA

/CRITERIA=PIN(.05) POUT(.10)

/NOORIGIN

/DEPENDENT happy

/METHOD=ENTER fe R1 rlgdgr.