

## Домашнее задание

### Deadline: 23.59 12 апреля 2021

*Общая постановка задачи* Задание выполняется на массиве [CDA\\_hw.sav](#). Источник данных – Pew Research Center. Тематика – восприятие степени угрозы коронавируса здоровью индивида. Описание переменных представлено ниже.

COVIDTHREAT\_b\_W63.5 - dependent variable

How much of a threat, if any, is the coronavirus outbreak for your personal health?

| Value | Label          |
|-------|----------------|
| 1     | A major threat |
| 2     | A minor threat |
| 3     | Not a threat   |
| 99    | No answer      |

F\_SEX - a person's biological sex

| Value | Label     |
|-------|-----------|
| 1     | male      |
| 2     | female    |
| 99    | No answer |

F\_AGECA1 age categories

| Value | Label        |
|-------|--------------|
| 1     | 18-29        |
| 2     | 30-49        |
| 3     | 50-64        |
| 4     | 65 and older |
| 99    | No answer    |

F\_EDUCAT2: education category

| Value | Label                           |
|-------|---------------------------------|
| 1     | Less than high school           |
| 2     | High school graduate            |
| 3     | Some college, no degree         |
| 4     | Associate's degree              |
| 5     | College graduate/some post grad |
| 6     | Postgraduate                    |
| 99    | No answer                       |

F\_INCOME Family income

| Value | Label                            |
|-------|----------------------------------|
| 1     | Less than \$10,000               |
| 2     | \$10,000 to less than \$20,000   |
| 3     | \$20,000 to less than \$30,000   |
| 4     | \$30,000 to less than \$40,000   |
| 5     | \$40,000 to less than \$50,000   |
| 6     | \$50,000 to less than \$75,000   |
| 7     | \$75,000 to less than \$100,000  |
| 8     | \$100,000 to less than \$150,000 |
| 9     | \$150,000 or more                |
| 99    | No answer                        |

**Задание 1.** Подготовьте данные к анализу.

1. Оставьте в массиве только вышеприведенные переменные.
2. Представьте все «No answer» как пропущенные значения.
3. Переверните шкалу зависимой переменной так, чтобы категории «3» соответствовало восприятие угрозы коронавируса здоровью индивида как серьезной.
4. Создайте для каждой возрастной категории дамми-переменную.

**Задание 2.** Оцените логит-модель порядковой регрессии. В качестве предикторов возьмите пол респондента, возрастные категории (как дамми-переменные, созданные на предыдущем шаге, используйте возрастную категорию 30 – 49 лет как базовую), уровень образования и доход семьи в качестве псевдо-интервальных переменных. Представьте интерпретацию оценок логит-модели в терминах отношения шансов (проиллюстрируйте на примере интерпретации двух оценок коэффициентов).

**Задание 3.** Рассчитайте дискретные изменения в вероятностях попасть в ту или иную категорию зависимой переменной при переходе от минимального уровня образования к максимальному. Рассмотрите мужчин возраста 18 – 29 лет со средним доходом. Проинтерпретируйте полученные значения (в том числе, и значимость).

**Примечание:** В проверочной работе также может встретиться интерпретация предельных эффектов и стандартизированных оценок коэффициентов.

**Задание 4.** Тестирование ограничений посредством теста Вальда

1. Протестируйте, является ли коэффициент при переменной «пол респондента» значимым.
2. Являются ли коэффициенты при переменных возрастных категорий одновременно незначимыми?

**Задание 5.** Постройте график зависимости предсказанной вероятности попасть в ту или иную категорию зависимой переменной (восприятия риска коронавируса для здоровья индивида) от дохода семьи для женщин 65 лет и старше с наиболее высоким уровнем образования. Проинтерпретируйте этот график.

**Задание 6.** Сделайте выводы о том, насколько хорошо модель предсказывает зависимую переменную, на основе мер Accuracy (сравнение с baseline accuracy), чувствительности и специфичности.

**Задание 7.** Протестируйте, справедливо ли применительно к модели допущение о параллельности регрессий. Сделайте предварительные выводы на основе сравнения оценок коэффициентов в соответствующих моделях с бинарным откликом.

**Задание 8.** Оцените glogit-модель без всяких ограничений и проинтерпретируйте полученные оценки. Сравните результаты с оценками из предыдущего задания.

**Задание 9.** Проверьте допущение о параллельности регрессий посредством формального теста. При необходимости оцените подходящую модель с частично-пропорциональными шансами.