

## Домашнее задание №7.

Логистическая регрессия и логлинейный анализ  
Группа 233**Дедлайн: 11 июня**

База данных `Titanic.sav` содержит данные, собранные после крушения “Титаника” в 1912 г. База данных включает в себя пять переменных: класс (первый, второй, третий, члены команды корабля), пол (мужчины, женщины), возраст (взрослые, дети), выживаемость (выжили или нет в результате катастрофы) и количество наблюдений в ячейках таблицы сопряженности перечисленных четырех переменных.

1. Проанализируйте двумерные таблицы сопряженности переменных класс и выживаемость, пол и выживаемость, возраст и выживаемость с помощью критерия хи-квадрат. Есть ли статистически значимая связь между этими парами переменных? У каких пассажиров была выше вероятность выжить в результате катастрофы: мужчин или женщин, пассажиров какого класса, детей или взрослых? Рассчитайте шансы выживания для различных категорий пассажиров и отношения шансов между категориями.
2. Оцените модель логистической регрессии, в которой выживаемость является зависимой переменной, а пол, возраст и класс — независимыми переменными. Добавьте в модель эффекты взаимодействия между различными парами независимых переменных и эффект взаимодействия трех переменных. (Возможно, вам понадобится оценить несколько моделей). Какие эффекты статистически значимы, а какие нет? Что это означает с содержательной точки зрения? Проинтерпретируйте результаты анализа.
3. Допустим, что все переменные в модели имеют равный статус (т.е. нет независимых и зависимых переменных). С помощью логлинейного анализа исследуйте связь между четырьмя категориальными переменными. Оцените модель независимости. Хорошо ли она описывает данные (подтвердите свой ответ соответствующими статистическими параметрами)? Оцените насыщенную модель. Что она означает? С помощью функции выбора наилучшей модели найдите модель, наилучшим образом описывающую данные. Какие эффекты взаимодействия она включает? Приведите статистические характеристики этой модели (хи-квадрат, отклонение, значимость). Проинтерпретируйте модель содержательно. Сделайте общие содержательные выводы по результатам логлинейного анализа.