

## Семинар. Модели бинарного выбора

**Задание 1.** Вероятность того, что студент с отличной оценкой по дисциплине, но без знания *LaTeX*, будет выбран ассистентом ( $Y = 1$ ) составляет 0.85. Вероятность того, что студент с отличной оценкой по дисциплине будет выбран ассистентом после освоения *LaTeX* составляет 0.9. Далее мы будем рассматривать только студентов, получивших оценку «отлично» за курс, так как это одно из основных условий, позволяющих претендовать на позицию ассистента.

1. Во сколько раз освоение *LaTeX* увеличивает вероятность быть выбранным ассистентом?
2. Рассчитайте шансы быть выбранным ассистентом без знания *LaTeX*
3. Рассчитайте, во сколько раз освоение *LaTeX* при прочих равных увеличивает шансы быть выбранным ассистентом
4. Ответьте на предыдущий вопрос, но в терминах изменения в процентах

### Задание 2.

Оценена логистическая модель, в которой предикторы  $x_{1i}$  — возраст (в годах),  $x_{2i}$  — пол (дамми-переменная: 0 — женщина, 1 — мужчина), зависимая переменная — курение (1 — курит, 0 — в противном случае).

1. Известно, что  $\hat{\beta}_0 = 0$ ,  $\hat{\beta}_1 = 0.01$  (оценка коэффициента при возрасте), и  $\hat{\beta}_2 = 0.5$  (оценка коэффициента при переменной «пол»). Запишите спецификацию модели и проинтерпретируйте эти исходные оценки коэффициентов в допущении об их статистической значимости.
2. Для указанной модели рассчитайте отношение шансов второго индивида к первому, если первый — женщина 25 лет, а второй — мужчина 25 лет. Проинтерпретируйте полученное значение.
3. Для указанной модели рассчитайте отношение шансов второго индивида к первому, если первый индивид — женщина 30 лет, а второй — мужчина 25 лет. Проинтерпретируйте полученное значение.
4. Для указанной модели рассчитайте отношение шансов второго индивида к первому, если первый — мужчина 20 лет, а второй — женщина 42 лет. Проинтерпретируйте полученное значение.
5. Преобразуйте исходные оценки коэффициентов при предикторах в отношения шансов и проинтерпретируйте указанные значения.

**Задание 3.** По данным города N. известно, что женщины доживают до 70 лет в 8 случаях из 10, а мужчины — в 4 случаях из 10. Рассчитайте отношение шансов дожить до 70 лет женщин — жительниц города N. — к мужчинам — жителям данного города. Проинтерпретируйте полученное значение.

**Задание 4.** Заказывающий в кафе ланч посетитель должен выбрать в качестве напитка чай или красное вино.

1. Для оценки вероятности того, что посетитель выберет красное вино (*красное вино закодировано как «1», чай — «0»*), используется логит-модель. Получены следующие оценки: константа равна 0.25, коэффициент при предикторе «основное блюдо» (*переменная принимает значение 1, если посетитель выбрал стейк из говядины, 0 — посетитель выбрал в качестве основного блюда рагу из овощей*) равен 1.36, коэффициент при предикторе «за рулем» (*1 — если посетителю в этот день еще предстоит вести машину, 0 — в противном случае*) составляет −2.4. Рассчитайте предсказанную вероятность выбора красного вина на ланч, если посетитель предпочел стейк из говядины и ему предстоит еще сесть за руль в этот день.

2. Аналогичные оценки в пробит-модели составляют 0.16 (константа), 0.7 («основное блюдо»), -1.45 («за рулем»). Рассчитайте теперь уже по оценкам пробит-модели предсказанную вероятность выбора красного вина на ланч, если посетитель предпочел стейк из говядины и ему предстоит еще сесть за руль в этот день.

3. Проинтерпретируйте оценки логит-модели в терминах отношения шансов.

**Задание 5.** Ниже представлена confusion matrix для порогового значения вероятности = 0.5

		Reference	
Prediction		0	1
0	430	192	
1	252	723	

Рассчитайте

1. Accuracy и Baseline accuracy
2. Ошибку первого рода
3. Ошибку второго рода
4. Мощность критерия
5. Меру специфичности (specificity)
6. Меру чувствительности (sensitivity)
7. Нарисуйте ROC-кривую для модели с точками в  $\tilde{p}=0.5$ ,  $\tilde{p}=0$  и  $\tilde{p}=1$ .

**Задание 6.** На основе ниже представленной выдачи рассчитайте статистику Hosmer-Lemeshow и проверьте нулевую гипотезу данного теста.

	y0	y1	yhat0	yhat1	total
[0.049,0.134]	798	81	783.5051	95.49489	879
(0.134,0.175]	735	141	740.6315	135.36848	876
(0.175,0.22]	681	194	702.4038	172.59618	875
(0.22,0.293]	659	218	655.0052	221.99485	877
(0.293,0.639]	568	309	559.4544	317.54561	877