

Дорогой храбрый воин или храбрая воительница! Удачи тебе на малом празднике по прикладной статистике! Начни с того, что напиши клятву и подпишись под ней:

Я клянусь честью студента, что буду выполнять эту работу самостоятельно.

А теперь — задачи:

1. Компания «ГолденАльп» тестирует два новых вкуса шоколада: с орешками и солёной карамелью. Фокус-группа разбивают на две непересекающиеся части: N_1 человек пробуют шоколад с орешками, а N_2 — с солёной карамелью.

Каждый участник пробует лишь один тип шоколада и одобряет или не одобряет опробованный вкус.

Пусть X_1 — число человек, одоббивших шоколад с орешками, а X_2 — одоббивших шоколад с солёной карамелью.

Будем предполагать, что $X_1 \sim \text{Bin}(N_1, p_1)$, $X_2 \sim \text{Bin}(N_2, p_2)$.

Руководство компании «Голден Альп» хочет узнать, есть ли основание полагать, что один вкус шоколада предпочитается другому.

Для этого её статистический отдел предлагает исследовать величину $p = p_1 - p_2$.

- а) Найди \hat{p}_{ML} .
 - б) Построй 95% доверительный интервал для p .
 - в) Подробно опиши, как построить 95% доверительный интервал для p при помощи какого-нибудь из методов бутстрэпа (метод выбирай сам).
 - г) По результатам эксперимента оказалось, что $N_1 = N_2 = 500$, $X_1 = 400$, $X_2 = 390$. Сформулируй гипотезу, которая позволит ответить на вопрос руководства компании. Протестируй эту гипотезу при помощи LR и LM тестов на уровне значимости 5%.
2. Пусть X и Y — n -мерные случайные векторы, образованные из независимых одинаково распределённых случайных величин.
Докажи, что $D_{KL}(p_X, p_Y) = nD_{KL}(p_{X_1}, p_{Y_1})$, где p — функция плотности.
 3. Докажи асимптотическую эквивалентность тестов LR и W в следующем смысле:

$$\text{plim}_{n \rightarrow \infty} \frac{W}{LR} = 1.$$

Не знаешь с чего начать? Разложи логарифм правдоподобия по Тейлору и найди приблизительное выражение для LR -статистики.

О храбрый воин и храбрая воительница! На следующей страничке есть ещё задачи!

4. Исходная выборка y — вектор из n независимых случайных величин, равновероятно принимающих значения 0 и 1. Пусть y^* — одна из бутстэп-выборок.
- а) Просто для удобства выпиши $E(y_i)$, $\text{Var}(y_i)$, $E(\bar{y})$, $\text{Var}(\bar{y})$.
 - б) Найди $E(y_i^*)$, $\text{Var}(y_i^*)$, $E(\bar{y}^*)$, $\text{Var}(\bar{y}^*)$.
 - в) Найди $\text{Cov}(y_i, y_i^*)$, $\text{Cov}(\bar{y}, \bar{y}^*)$.
5. У меня есть три монетки. Они выпадают орлом с вероятностями p_1 , p_2 и $p_3 = 1$. Я провожу эксперимент из 100 раундов.
- В каждом раунде я равновероятно выбираю одну из монеток. Подбрасываю её два раза и записываю число выпавших орлов.
- После окончания эксперимента у меня остаётся на бумажке 100 записанных чисел. Какая монетка подкидывалась в каждом раунде, я не помню.
- Опиши ЕМ-алгоритм для оценивания неизвестных p_1 и p_2 .
- Если формулы для какого-то шага выводятся в явном виде, то выведи их. Если формулы для какого-то шага не выводятся в явном виде, то объясни, какая оптимизационная задача будет решаться численно.
6. Рассмотрим модель множественной регрессии $y = X\beta + u$, оцениваемую при помощи МНК. Число наблюдений равно $n = 500$, число регрессоров равно $k = 10$, включая константный. Все регрессоры ортогональны друг другу.
- а) Мелиодас строит оценку МНК по константному и первым следующим за ним четырём регрессорам, а Элизабет строит оценку МНК по константному и оставшимся пяти регрессорам. Покажи на *единой* картинке МНК \hat{y} , TSS , ESS , RSS и R^2 в их регрессиях.
 - б) Эсканор строит оценку МНК по всем 10 регрессорам. Покажи на картинке МНК из предыдущего пункта \hat{y} , TSS , ESS , RSS и R^2 в его регрессии.
 - в) Как соотносятся R^2 в регрессиях Эскарнора, Мелиодаса и Элизабет?