

Коновалова Александра Сергеевна (kas-tstu@yandex.ru)

Как я использовала бутстрэп для проверки решений Богдана*?

К рекомендованным пакетам подгрузила *scipy.stats* для работы с распределением Бернулли, ведь я не волшебник, я только учусь...

```
import scipy.stats as sts
```

Основываясь на *результатах опроса 100 жителей мегаполиса*, половина из которых готова поддержать кандидатуру ученого кота на выборах председателя академии наук (см. приложение), оценила **доверительные интервалы для вероятности того, что случайный житель проголосует за ученого кота** (значит, и для доли людей, которые за него проголосуют):

```
x=sts.bernoulli(0.5).rvs(100)

boot_x = IIDBootstrap(x, seed=111111)
boot_x.conf_int(np.mean, method='percentile', reps=10000, size=0.95)
```

Получила `array([[0.41], [0.61]])`, т.е. ответ (до одного знака после запятой) совпал с решением Богдана, и у ученого кота как минимум есть шансы выйти во второй тур, а при везении и победить!

Причем квантили `[np.quantile(x, 0.025), np.quantile(x, 0.975)]` не прокатили 😊

Далее, учитывая *результаты испытаний* изобретенного ученым котом и Лаврентием (независимо друг от друга) лекарства от лени (см. приложение), нашла **95%-й доверительный интервал для разницы успеха** (вероятности успешного действия) изобретенных лекарственных препаратов:

```
x1=sts.bernoulli(0.8).rvs(225)
x2=sts.bernoulli(0.6).rvs(100)

def mean_diff(x, y):
    return np.mean(x) - np.mean(y)

boot_xy = IndependentSamplesBootstrap(x1, x2, seed=111111)
boot_xy.conf_int(mean_diff, reps=10000, size=0.95, method='basic')
```

Получила `array([[0.11888889], [0.34111111]])`, что вновь (до одного знака после запятой) совпало с решением Богдана, и убедилась, что ученый кот, действительно, заслуживает победы на выборах председателя академии наук!

* С искренним и глубоким уважением к **Богдану Потанину**, автору курса [http://wiki.cs.hse.ru/Теория вероятностей и статистика, МИРЭК, 2023-2024](http://wiki.cs.hse.ru/Теория_вероятностей_и_статистика,_МИРЭК,_2023-2024)

Асимптотические доверительные интервалы для доли

Доля

- Рассмотрим выборку $X = (X_1, \dots, X_n)$ из распределения Бернулли с параметром $p \in (0, 1)$.
- Используя теоремы Муавра–Лапласа и Слущкого получаем $100(1 - \gamma)$ процентный асимптотический доверительный интервал для p :

$$\left[\bar{X}_n - z_{1-\gamma/2} \sqrt{\frac{\bar{X}_n (1 - \bar{X}_n)}{n}}, \bar{X}_n + z_{1-\gamma/2} \sqrt{\frac{\bar{X}_n (1 - \bar{X}_n)}{n}} \right]$$

Пример: по результатам опроса 100 жителей очень большого города оказалось, что половина из них готова поддержать на выборах председателя академии наук кандидатуру ученого кота. Найдем реализацию 95%-го асимптотического доверительного интервала для вероятности того, что случайно выбранный житель проголосует за ученого кота (исходя из ЗБЧ она будет приблизительно равняться доле людей, которые за него проголосуют). Поскольку $\bar{x}_{100} = 50/100 = 0.5$ и $z_{0.975} \approx 1.96$, то:

$$\left[0.5 - 1.96 \sqrt{\frac{0.5(1 - 0.5)}{100}}, 0.5 + 1.96 \sqrt{\frac{0.5(1 - 0.5)}{100}} \right] = [0.402, 0.598]$$

Асимптотические доверительные интервалы для доли

Разница долей

- Рассмотрим независимые выборки $X = (X_1, \dots, X_n)$ и $Y = (Y_1, \dots, Y_m)$ из распределений Бернулли с параметрами $p_X \in (0, 1)$ и $p_Y \in (0, 1)$ соответственно.
- Используя теоремы Муавра–Лапласа и Слущкого получаем $100(1 - \gamma)$ процентный асимптотический доверительный интервал для $p_X - p_Y$:

$$\left[\bar{X}_n - \bar{Y}_m - z_{1-\gamma/2} \sqrt{\frac{\bar{X}_n (1 - \bar{X}_n)}{n} + \frac{\bar{Y}_m (1 - \bar{Y}_m)}{m}}, \bar{X}_n - \bar{Y}_m + z_{1-\gamma/2} \sqrt{\frac{\bar{X}_n (1 - \bar{X}_n)}{n} + \frac{\bar{Y}_m (1 - \bar{Y}_m)}{m}} \right]$$

Пример: ученый кот и Лаврентий независимо друг от друга изобрели лекарство от лени. Ученый кот испытал свое лекарство на 225 добровольцах, а Лаврентий – на 100. Среди добровольцев ученого кота лениться меньше стали 180 испытуемых, а у Лаврентия – 60%. Найдите реализацию 95%-го асимптотического доверительного интервала для разницы в вероятностях успешного действия лекарства ученого кота и Лаврентия. Обратим внимание, что $\bar{x}_{225} = 180/225 = 0.8$, $\bar{y}_{100} = 0.6$ и $z_{0.975} \approx 1.96$, поэтому:

$$\left[0.8 - 0.6 - 1.96 \sqrt{\frac{0.8(1 - 0.8)}{225} + \frac{0.6(1 - 0.6)}{100}}, 0.8 - 0.6 + 1.96 \sqrt{\frac{0.8(1 - 0.8)}{225} + \frac{0.6(1 - 0.6)}{100}} \right] \approx [0.09, 0.31]$$

Ист.: [http://wiki.cs.hse.ru/Теория вероятностей и статистика](http://wiki.cs.hse.ru/Теория_вероятностей_и_статистика), МИРЭК, 2023-2024