

Как я использовала бутстрэп?

Источник данных - датасет Heart Disease, содержащий записи, связанные с диагностикой сердечных заболеваний у пациентов. Датасет содержит 918 записей (демография пациентов, показатели жизнедеятельности, симптомы и потенциальные факторы риска, связанные с заболеваниями сердца). Первые 10 записей:

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
0	40	M	ATA	140	289	0	Normal	172	N	0.0	Up	0
1	49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
2	37	M	ATA	130	283	0	ST	98	N	0.0	Up	0
3	48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
4	54	M	NAP	150	195	0	Normal	122	N	0.0	Up	0
5	39	M	NAP	120	339	0	Normal	170	N	0.0	Up	0
6	45	F	ATA	130	237	0	Normal	170	N	0.0	Up	0
7	54	M	ATA	110	208	0	Normal	142	N	0.0	Up	0
8	37	M	ASY	140	207	0	Normal	130	Y	1.5	Flat	1
9	48	F	ATA	120	284	0	Normal	120	N	0.0	Up	0

Оцениваемый параметр - математическое ожидание величины артериального давления в состоянии покоя (Resting BP). Точечная оценка $\mu_{\text{mean}}(bp) = 132.40$. Создание бутстрэпа и первые 10 значений:

```
n_boot_bp = 10000
mu_hat_star_bp = [np.mean(rng.choice(bp, len(bp))) for i in range(n_boot_bp)]
mu_hat_star_bp[1:10]
```

```
[132.57407407407408,
 131.9281045751634,
 132.6165577342048,
 133.05228758169935,
 131.6949891067538,
 132.640522875817,
 132.36056644880173,
 131.79520697167757,
 132.7450980392157]
```

Создание 95% доверительного интервала:

```
[np.quantile(mu_hat_star_bp, 0.025), np.quantile(mu_hat_star_bp, 0.975)]
[131.23090958605664, 133.6056917211329]
```

Выводы: с помощью наивного бутстрэпа получена интервальная оценка математического ожидания величины артериального давления в состоянии покоя (Resting BP). Точечная оценка, равная 132.40, вошла в 95% доверительный интервал [131.23; 133.61].