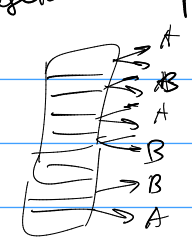


Привет

AB - эксперимент
AA - эксперимент

чтобы сделать эксперимент не нужно
A старое
B новое
YB - YA
чтобы проверить, всё ли ок.



Бутстреп - самобыто методов.

group _i	y _i
A	200
A	250
B	300
A	260
B	310
B	240

$$P(\bar{y}_B^{boot} - \bar{y}_A^{boot} \geq 46.6) = \frac{1+1+1}{20} = \frac{3}{20} = 0.15$$

A = "красивое" ← старое приложение
B = "хорошее утро" ← новая техника

!!! учтите порядок ранжирования!!!

$$C_6^3 = \frac{6!}{3!3!} = 6 \cdot 5 \cdot 4 / 6 = 20$$

$\bar{y}_B - \bar{y}_A \approx 46.6$

фактическое
оказывается ли такая величина
сигн. или факторами?

не знаем точно!

Итак, предположим, что $(y_i | group_i = A) \sim (y_i | group_i = B)$

⇒ если я знаю сигн. по представлю буквы A и B, то закон распределения YB - YA совпадает.

Рассмотрим эти случаи переоткрытия

A	200
A	250
A	300
B	260
B	310
B	240

A	200
A	250
B	300
B	260
A	310
B	240

маленький тест
* $y_i \sim N(\mu, \sigma^2)$
* $n \gg 0$

в бутстрепе подходе
предполагаем, что

$C_6^3 = 20$
каждого
переоткрытия

бутстрепирование

$\bar{y}_B - \bar{y}_A = \frac{60}{3} = 20$ $\bar{y}_B - \bar{y}_A = \frac{40}{3}$

бутстреп - повторение = для бутстрапа $\bar{y}_B - \bar{y}_A$
более правдоподобно
или наоборот
→ $\bar{y}_B - \bar{y}_A$ бутстрепирование

Y_{exp}

или хотим $\bar{y}_{boot} \approx \bar{y}$

модели посчитал

$$P(\bar{y}_{boot} \leq 346) \approx P(\bar{y} \leq 346)$$

исх. дан
выборки

y ₁	340
y ₂	370
y ₃	360
y ₄	340
y ₅	320

$y_i \sim F$
↑
кажд.
зона
распр.

$$F_{y_i}(u) = \begin{cases} 0 & \text{если } u < y_i \\ P(y_i \leq u) & \text{иначе} \end{cases}$$

$\bar{y} = 346$

и вы же понимаете, что это "исходное взв-ое среднее"

исходное
не
равно
исходному

дуга с рн - взв-ка n1

y ₁ [*]	360
y ₂ [*]	340
y ₃ [*]	340
y ₄ [*]	340
y ₅ [*]	320

$y_i^* \sim F^*$

$\rightarrow \bar{y}^* = 340$

взв-ка y-распр, взв-ка р-на распр.

F^* :

340	370	360	340	320
45	15	45	45	15

дуга с рн - взв-ка n2

y ₁ [*]	340
y ₂ [*]	340
y ₃ [*]	340
y ₄ [*]	340
y ₅ [*]	340

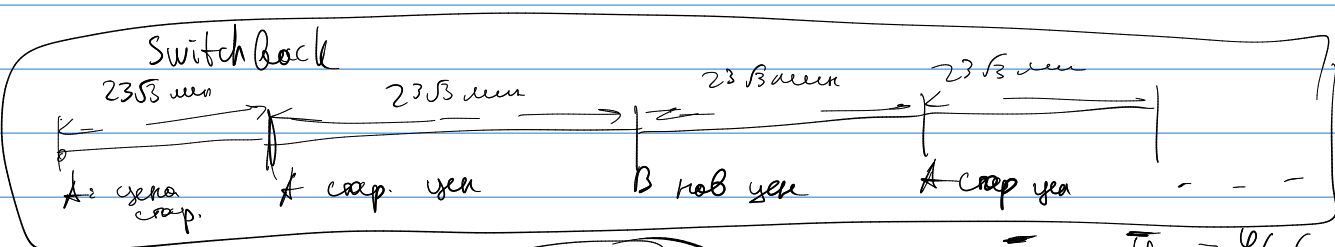
$\rightarrow \bar{y}^* = 340$

y_i^* взв-ка равнов-ко у всех
имеющихся y_1, \dots, y_n (исх-х)

$$Var(\bar{y}) = \frac{Var(y_i)}{n}$$

$n_{boot} = 10000$: $\bar{y}^*, \bar{y}^*, \dots, \bar{y}^*$ 10000 раз

дуга с рн R значение = доля \bar{y}^* , ок-ся меньше \bar{y} .



$\bar{y}_B^* - \bar{y}_A^* = \dots$
 $\bar{y}_D^* - \bar{y}_A^* = \dots$

$\bar{y}_B - \bar{y}_A = 46.6$

Упр.

У нас - 1000 наблюдений $n = 1000$ наблюдений.

На базе этой выборки строим другую выборку.

$N_1, N_2, N_3, \dots, N_{1000}$ - сколько раз то или иное наблюдение было скопировано в эту выборку.

a) $N_1 + N_2 + N_3 + \dots + N_{1000} = 1000$

b) $N_i \sim \text{Bin}(n=1000, p=\frac{1}{1000})$

в) $N_i \sim ?$ $\text{Pois}(\lambda=1)$ $N(??)$

г) $E(N_i) = ?$ $\text{Var}(N_i) = ?$

д) $\text{Cov}(N_1, N_2) = ?$