

$$P(N=2) = \frac{1}{60} \cdot \frac{1}{60} \cdot (\frac{59}{60})^{58} \cdot C_{60}^{2}$$

$$N \sim \text{kinoh} \left( N_1 p \right) = \text{kin} \left( 60, \frac{1}{60} \right)$$

$$p(N=k) = C^k \cdot p^k \cdot (1-p)^{kk}$$

$$P(N \in \Delta) = P(N=0) + P(N=1)$$

I have a sample  $X_1, \ldots, X_{60}$ . I generate one naive bootstrap sample  $X_1^*, \ldots, X_{60}^*$ . Let L be the number of initial observations missing in the bootstrap sample. Find the expected value  $\mathbb{E}(L)$ .

$$e_i = \frac{1}{0}$$
, missing  $\frac{59}{60}$  not missing  $\frac{1}{60}$ 

$$E(e_i) = p = \frac{59}{60}$$
  $Van(e_i) = pq = \frac{59}{60} \cdot \frac{1}{60}$ 

$$\pm (L) = 1 \cdot \left(\frac{59}{60}\right)^{60} + 1 \cdot \left(\frac{59}{60}\right)^{60} + \dots + 1 \cdot \left(\frac{59}{60}\right)^{60}$$

$$= 60 \cdot \left(\frac{59}{60}\right)^{60}$$

We have data of an AB experiment:  $\bar{X}_a = 5.1$ ,  $\bar{X}_b = 5.6$ ,  $n_a = 18$ ,  $n_b = 12$ ,  $\sum (X_i^a - \bar{X}_a)^2 = 840$ ,  $\sum (X_i^b - \bar{X}_b)^2 = 820$ .

Calculate the estimate of variance of  $\bar{X}_a - \bar{X}_b$  for the Welch test.

V

Welch Test

Ho:

Ma=Mb

Ass.

6a \$ 66

(V - V.)

xa - xb - 0 ~ ts

(h,-1) S1 + (h2-1) · S2

| h, The

= 1 TSS,

$$\frac{2}{S_2} = \frac{7SS_2}{N-1}$$

$$Van(\overline{X}_{a}-\overline{X}_{b}) = \sqrt{\frac{S_{a}^{2} + \frac{S_{b}^{2}}{h_{a}}}{h_{a}}} = \sqrt{\frac{840/17}{18} + \frac{820/11}{12}}$$

### 3. Problem

We have data of an AB experiment:  $\bar{X}_a = 5.1$ ,  $\bar{X}_b = 5.6$ ,  $n_a = 18$ ,  $n_b = 12$ ,  $\sum (X_i^a - \bar{X}_a)^2 = 840$ ,  $\sum (X_i^b - \bar{X}_b)^2 = 820$ .

Calculate the estimate of variance of  $\bar{X}_a - \bar{X}_b$  for the Welch test.

X:~ N(5,3) 3=62 ZX: Variables  $X_1, X_2, ..., X_{20}$  are independent and normally distributed  $\mathcal{N}(5;3)$ .

Calculate the variance

$$\operatorname{Var}\left(\sum_{i=1}^{20}(X_i-\bar{X})^2\right)$$
. Cor  $\left(X_i-\bar{X},X_j-\overline{X}\right)\neq 0$ 

$$\chi_{\mu}^{2} = \sum_{i=1}^{\kappa} g_{i}^{2}, \quad g_{i} \sim N(\alpha_{i})$$

$$\sum \frac{\left(X_{i}-\overline{X}\right)^{2}}{6^{2}}=\sum g^{2}$$

$$S^2 = \frac{1}{h-1} \cdot TSS$$

$$\rightarrow k \cdot k^2 \sim \chi^2$$

$$\frac{(h-1) S^2}{6^2} \sim \chi_{h-1}^2$$

$$L_{3} F = \frac{Var(\cdot)}{Var(\cdot)} \sim \frac{\chi^{2}}{\chi^{2}} = F_{p,q}$$

$$Var\left(\frac{h-1}{6^2} S^2\right) = Var\left(\chi_{h-1}^2\right)$$

$$\frac{(N-1)^2}{6^4} \cdot \text{Var}\left(\frac{755}{h-1}\right) = \text{Var}\left(\frac{72}{h-1}\right)$$

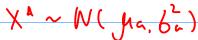
$$Van(X^{2}) = Van(g^{2}) =$$

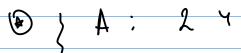
$$E(g^{4}) - E^{2}(g^{2}) = \mu_{4} - \mu_{2}^{2} =$$

$$Van\left(\chi_{k}\right) = Van\left(\sum_{j=1}^{k} \sum_{i=1}^{k} Van\left(s_{i}^{2}\right) = \sum_{j=1}^{k} Van\left(s_{i}^{2}\right)$$

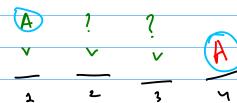
I have results of two runners A and B: 2 results for A and 3 results for B. The running time for both runners are continuously distributed and their distribution are equal.

What is the probability that the maximal rank of running times of the runner A will be equal to 4?





$$rank(X_{5}^{b}) = 5$$



$$P(\max (2anh A) = 4) = \frac{C_1^2}{C_2^5}$$

renh=; ?!ABB

AABBB

I have results of two runners A and B: 2 results for A and 3 results for B. The running times for both runners are continuously distributed.

Consider the Mann-Whitney test statistic  $U_a$  that positively depends on the rank sum for the runner A.

Find the expected value  $\mathbb{E}(A)$  under the pull-hypothesis of equal running time distributions.

$$V \sim \mathcal{N}\left(\frac{n_1 n_2}{2}; \frac{n_1 n_2 (n_1 + n_2 - 1)}{12}\right)$$

$$\frac{1 - \frac{2.3}{2}}{\sqrt{\frac{2.3.4}{12}}} = \frac{-2}{\sqrt{2}} \approx -1.4 \quad (-1.4) < 1.96$$

$$= 7 \text{ Ho is not ney.}$$

Consider three variables: target variable  $y_i$ , predictor  $x_i$  and indicator of treatment  $z_i \in \{0,1\}$ .

The treatment  $z_i$  was designed to be independent of  $x_i$ , but in fact  $x_i = f_i \cdot (1 + 0.08z_i)$ .

We suppose that  $z_i$  are Bernoulli with p = 0.4,  $f_i \sim \mathcal{N}(900; 9)$  and they are independent.

Find the probability limit

$$p\lim \frac{\sum (x_i - \bar{x})(z_i - \bar{z})}{n-1}.$$

Plin 
$$\frac{\overline{X}}{h \to \infty} = E(X) = \mu$$

$$\frac{\sum (X_i - \overline{X})^2}{h} = \lim_{h \to \infty} \frac{\sum (X_i - \overline{X})}{h - 1} = Van(X)$$

### 7. Problem

Consider three variables: target variable  $y_i$ , predictor  $x_i$  and indicator of treatment  $z_i \in \{0,1\}$ .

The treatment  $z_i$  was designed to be independent of  $x_i$ , but in fact  $x_i = f_i \cdot (1 + 0.08z_i)$ .

We suppose that  $z_i$  are Bernoulli with p = 0.4,  $f_i \sim \mathcal{N}(900; 9)$  and they are independent.

Find the probability limit

$$p\lim \frac{\sum (x_i - \bar{x})(z_i - \bar{z})}{n-1}.$$

$$\frac{p | \text{Im} \ Cov(X_{i}, Z_{i}) = Cov(X_{i}, (1+0,08,2), z_{i})}{n + \infty} = Cov(X_{i}, Z_{i}) = Cov(X_{i}$$

Consider three variables: target variable  $y_i$ , prediction  $x_i$  and indicator of treatment  $z_i \in \{0,1\}$ . The treatment  $z_i$  was assigned independently of  $x_i$ , total n = 100.

The sample covariance matrix C is provided. The order of variables is y, x and z. For example the sample covariance of x and z is 0.26.

$$C = \begin{pmatrix} 7.09 & 2.64 & 0.75 \\ 2.64 & 1.17 & 0.26 \\ 0.75 & 0.26 & 0.25 \end{pmatrix}$$

Consider CUPED with first regression given by  $\hat{y}_i = \hat{\alpha}_1 + \hat{\alpha}_2 x_i$  with residuals  $r_i = y_i - \hat{y}_i$ .

What is the sample covariance of r and z?

Cuped: 1) Y cuped = 
$$y - \theta x$$

$$\hat{\theta} = \frac{\hat{\omega}(x,y)}{\hat{\omega}(x,y)} \iff \hat{\beta} \text{ in } y(x)$$

2) 
$$Var(\overline{y}_{uped}) = (1-p^2) Var(\overline{y}) \leq Var(\overline{y})$$

$$\int_{X,Y}^{2} = R^{2}$$

$$= C\hat{O}(Y - \hat{O}(X, 2) =$$

$$Cov(y,z) - \frac{Cov(x,y)}{Van(x)} Cov(x,z) =$$