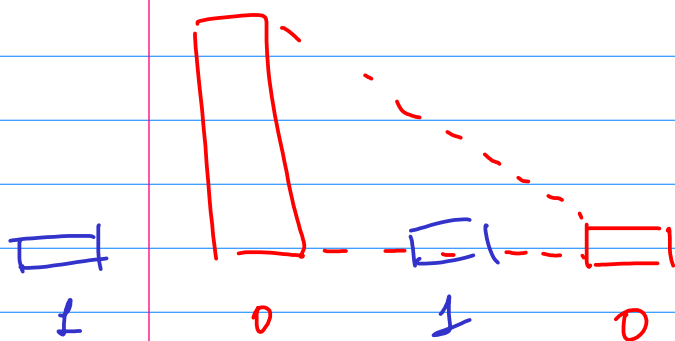


- convenience sampling
 - judgemental sampling
 - quota sampling
 - snowball sampling
- simple random sampling
 - systematic sampling
 - stratified sampling
 - cluster sampling

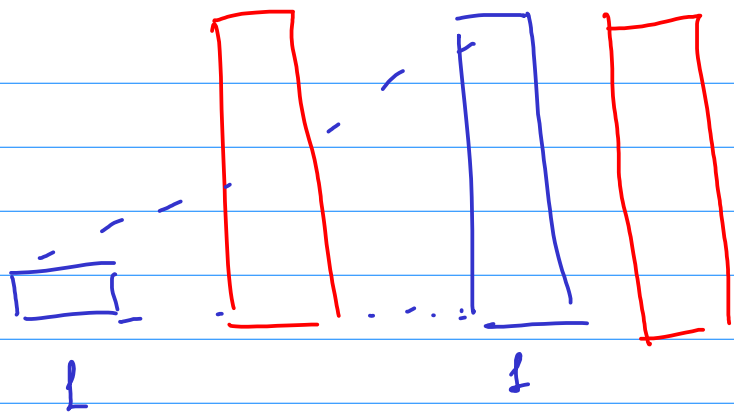
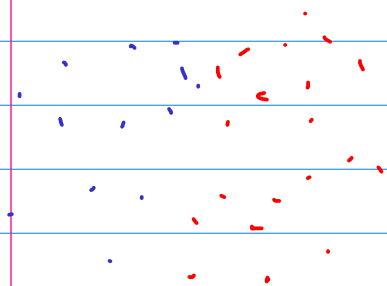
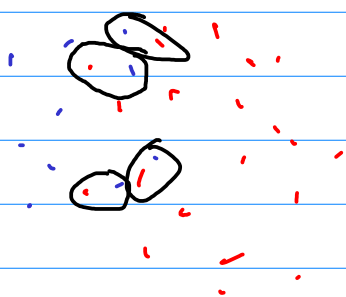
$$p_0 = p_1 = 0.5$$

Resampling



under-sampling

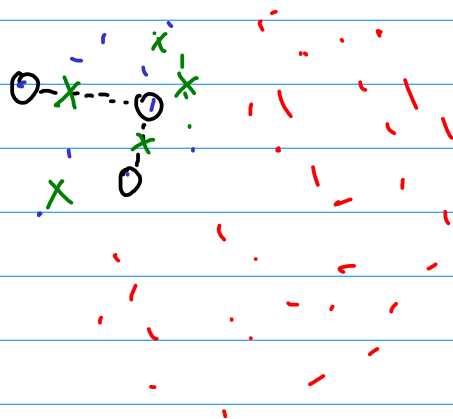
Tomek links



over-sampling

SMOTE

(Synthetic minority
over-sampling)



Multiple hyp. testing

H_0 : no relationship

H_a : there is relationship

20 hypothesis $\alpha = 5\%$

$P(\text{one significant result} | H_0) =$

$1 - P(\text{no sig. results} | H_0) =$

$$1 - (1 - \alpha)^m = 1 - 0,95^{20} = 0,64$$

$$\text{FWER} \leq \alpha$$

$$\text{FWER} = P_2(V \geq 1) = 1 - P(V=0)$$

V - # incorrectly
rej. hyp.

- Control FWER

1) Bonferroni correction

$$\alpha_i = \alpha / m$$

2) Holm - Bonferroni correction

$$\alpha_i = \frac{\alpha}{m + 1 - k}$$

k - rank of hyp.
by p-value

3) Sidak

$$\text{FWER} \leq \alpha$$

$$P(V \leq 1) = 1 - P(V = 0) \leq 1 - (1 - \alpha_1)^m = \alpha$$

$$\alpha_1 = 1 - (1 - \alpha)^{1/m}$$

4) Sidak - Holmes

$$\alpha_i = 1 - (1 - \alpha)^{1/m}$$

$$\alpha_i = 1 - (1 - \alpha)^{\alpha/m - l + 1}$$

$$\alpha_m = \alpha$$

if hyp. are ind.

$$\Rightarrow FWER \leq \alpha$$

if hyp. are jointly ind

\Rightarrow highest power

$$FPR = E\left(\frac{V}{R} \mid R > 0\right)$$

$$FPR \leq FWER$$

1) Benjamin - Hochberg

$$\alpha_i = k/m \cdot \alpha$$

control FDR at $\frac{m_0}{m} \alpha$,

where m_0 is #
of true H_0