SRSWR

simple random sampling
WITH
replacement

SRSWOR

without

N individuals
(population)

You're measuring some trait of these individuals

We'd like to understand the distribution of this trait in the popln

D/R example $\left\{ \underbrace{1, 2, \ldots, N}_{\text{individuals}} \right.$

i^{th} individual is 1 or 0 accordingly as D or R.

Sample: subset of selected individuals from Popln. on whom the trait will be measured

---

In order to collect a representative sample, we resort to randomness.

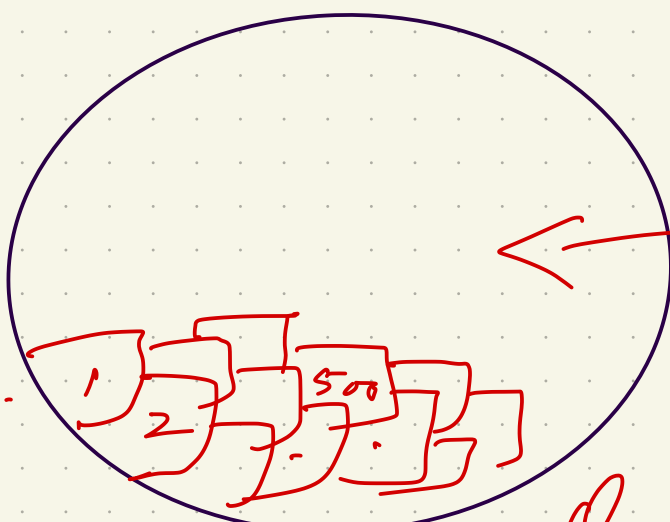
SRSWR

$1, 2, -\ -, N$

$z_1, z_2, \dots, z_N$: values of the trait.

$n << N$

Select one ticket at random. Record number. Say this is i_1.



N paper tickets

Query indiv. i_1 and get z_{i_1}

Return ticket i_1 back into globe.

Then repeat to get $i_2 \to$ record i_2.

Repeat n times.

$\{ i_1, i_2, \ldots, i_n \}$: Sample.

random quantities.

$$P(i_1 = m) = \frac{1}{N}$$
$1 \leq m \leq N$

$$P(i_k = m) = \frac{1}{N}$$

These follow a discrete uniform distribution.

$$P(i_1 = m_1, i_2 = m_2, \ldots, i_n = m_n)$$

$m_1, m_2, \ldots, m_n$ are fixed numbers

$\{ i_1 = m_1 \}, \ldots, \{ i_n = m_n \}$.

should be independent.

$$P(i_1 = m_1, i_2 = m_2, \ldots, i_n = m_n) = \prod_{k=1}^{n} P(i_k = m_k)$$

Random variables of interest:

$X_1, X_2, \ldots, X_n$ where

$X_\ell = Z_{i_\ell}$.

X_ℓ assumes values 1 or 0.

$P(X_\ell = 1) = P$ (a D is picked at ℓ'th draw)

parameter of interest

$$= \frac{N_D}{N} \equiv p_D.$$

where N_D = number of D's in popln.

X_ℓ is a Bernoulli random variable

Because the events associated with the i_ℓ's are independent, events associated with X_ℓ's are also independent.

$P(X_1 = \varepsilon_1, X_2 = \varepsilon_2, \ldots, X_n = \varepsilon_n)$

Where $\varepsilon_1, \varepsilon_2, \ldots$ is a fixed sequence of 0's and 1's.

$$= \prod_{i=1}^{n} P(X_i = \varepsilon_i)$$

$$P(X_i = \varepsilon_i) = \begin{cases} p_D & \text{if } \varepsilon_i = 1 \\ 1 - p_D & \text{if } \varepsilon_i = 0 \end{cases}$$

$$= p_D^{\varepsilon_i} (1-p_D)^{1-\varepsilon_i}$$

$$= p_D^{\sum \varepsilon_i} (1-p_D)^{n - \sum \varepsilon_i} \longrightarrow$$ joint prob. mass function of $(X_1, \ldots, X_n)$.

$\sum_{i=1}^{n} \varepsilon_i = \#$ of 1's in the sequence $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$.

How many possible values can the random n tuple $(X_1, \ldots, X_n)$ take?

2^n = size of all possible $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$.

$$P(X_3 = 1 \mid \underline{X_1 = 0,\ X_2 = 1})$$

$$= P(X_3 = 1) = p_D$$

$$\frac{P(\underline{X_1 = 0,\ X_2 = 1},\ X_3 = 1)}{P(X_1 = 0,\ X_2 = 1)}$$

$$= \frac{\cancel{P(X_1 = 0,\ X_2 = 1)}\, P(X_3 = 1)}{\cancel{P(X_1 = 0,\ X_2 = 1)}}$$

$$\sum_{i=1}^{n} X_i = \underline{S_n} \sim \text{Bin}(n, p)$$

follows

$$E S_n = np, \text{ or } E\left(\frac{S_n}{n}\right) = \underline{p_D}$$

proportion of D's in sample, denoted $\underline{\hat{p}}$ (estimate)

$$E(S_n) = E(X_1 + X_2 + \cdots + X_n)$$

$$= \underline{EX_1 + EX_2 + \cdots + EX_n}$$

uses additivity of expectation.

$$= n p_D$$

$$X_\ell \left\{ \begin{array}{l} P(X_\ell = 1) = p_D \\ P(X_\ell = 0) = 1 - p_D \end{array} \right.$$

$\underline{\text{so } \mathbb{E}(X_\ell) = p_D}$

for any ℓ

E(sum of random variables)

= sum of expectations of the random variables.

Follows without too much difficulty from our characterization of $\mathbb{E}$.

$\underline{\mathbb{E}(X+Y)}$: $\longrightarrow$ consider

$X \geq 0, Y \geq 0\}$

$$\mathbb{E}X = \lim_n \mathbb{E}(L_n)$$

where $L_1 \le L_2 \le L_3 \le L_4 \le \cdots$

L_i's are simple and L_i's $\uparrow X$.

Similarly $\mathbb{E}Y = \lim \mathbb{E}(\tilde{L}_n)$

$\tilde{L}_1 \le \tilde{L}_2 \le \cdots$, and $\tilde{L}_i$'s $\uparrow Y$.

look at $L_n + \tilde{L}_n$. Clearly it's also increasing, and $L_n + \tilde{L}_n \uparrow Y$

simple function.

$$\mathbb{E}(X+Y) = \lim_n \mathbb{E}(L_n + \tilde{L}_n)$$

$$= \lim_n (\mathbb{E}L_n + \mathbb{E}\tilde{L}_n)$$

$$= \lim_n \mathbb{E}L_n + \lim_n \mathbb{E}\tilde{L}_n$$

$$= \mathbb{E}X + \mathbb{E}Y.$$

We did this for nng r.v's but in general, we $X = X^+ - X^-$ trick.

$$\mathbb{E}(\alpha X) = \alpha \cdot \mathbb{E}(X).$$

So in general:

$$\boxed{\mathbb{E}\left[\sum_{i=1}^{p} \beta_i W_i\right] = \sum_{i=1}^{p} \beta_i \, \mathbb{E}(W_i)}$$

numbers

$\boxed{\text{SRSWOR}}$

without

Pick i_1. Record z_{i_1}.

Don't return i_1 back to globe.

Then sample i_2. Record z_{i_2}.

Discard.

Proceed till stage n.

$(i_1, i_2, i_3, \ldots, i_n)$.

$(X_1, X_2, \ldots, X_n)$

$$\boxed{X_\ell = z_{i_\ell}}$$

(a) What is the distribution of X_ℓ?

$1 \le \ell \le n$

(b) Are the X_ℓ's independent.

(c) What is the distribution of S_n?

$$S_n = X_1 + \cdots + X_n$$

(d) What is $E S_n$?

<u>Question (a)</u>: Distribution of X_1.

$$X_1 \sim \text{Bernoulli}(p_D)$$

$$P(X_1 = 1) = p_D.$$

$$X_2 \sim ?$$

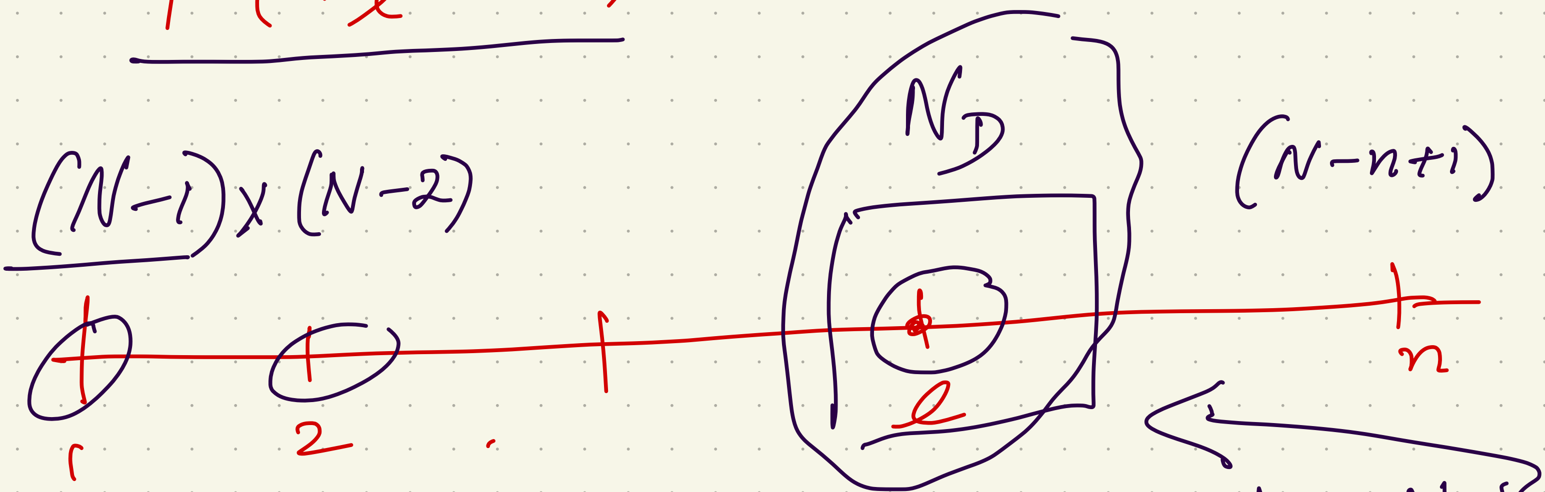
$*$ $$(1-p_D)\frac{N_D}{N-1} + p_D \frac{N_D - 1}{N-1} = p_D$$

$$P(X_2 = 1) = P(X_1 = 0, X_2 = 1) + P(X_1 = 1, X_2 = 1)$$

$*$

$$= \underline{P(X_1 = 0)\, P(X_2 = 1 \mid} X_1 = 0) + P(X_1 = \underline{1)\, P(X_2 = 1 \mid} X_1 = 1) \; *$$

$P(X_\ell = 1)$



Total no. of sequences of tickets possible?

$$^{N}P_n = N(N-1)\cdots(N-n+1).$$

→ all permutations are equally likely

Out of these, how many permutations give me $X_\ell = 1$?

$$N_D \times {}^{N-1}P_{n-1}$$

$$= N_D \times (N-1)\cdots(N-n+1)$$

$$P(X_\ell = 1) = \frac{N_D \times \cancel{(N-1)\times\cdots\times(N-n+1)}}{N \times \cancel{(N-1)\times\cdots\times(N-n+1)}}$$

$$= \frac{N_D}{N} = p_D.$$

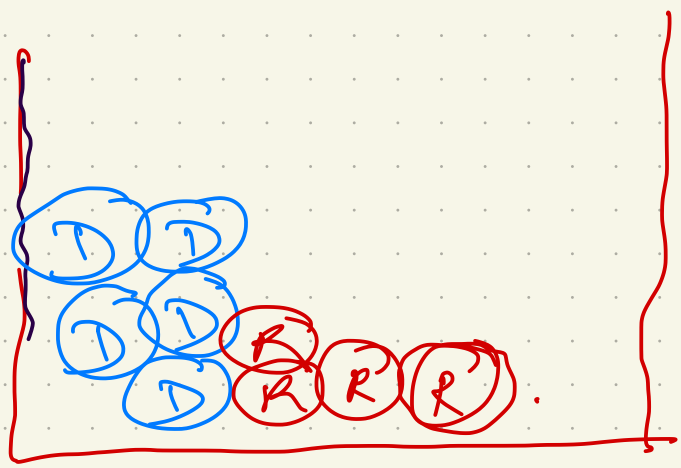
Notice ~~that~~ X_i's are not-independent

$$P(X_1 = 1, X_2 = 1) = p_D \cdot \frac{N_D - 1}{N - 1}$$

$$= p_D \frac{p_D - \frac{1}{N}}{1 - \frac{1}{N}}$$

What is the probability ~~that~~ $P(S_n = m)$

prob. ~~that~~ number of democrats in my sample of size n equals m?



Urn

N_D blue balls

$N - N_D$ red balls

Pick n balls from urn

P(I get m blue balls)

N balls

Picking n.

# of distinct groups: $\binom{N}{n}$

# of distinct groups with m D's and $n-m$ R's:

$$\binom{N_D}{m} \times \binom{N-N_D}{n-m}$$

P(picking m democrats in sample of size n)

$$P(V=m) = \frac{\binom{N_D}{m} \cdot \binom{N-N_D}{n-m}}{\binom{N}{n}}$$

random number of D's picked

$V \sim$ Hypergeometric dist. (N, N_D, n)