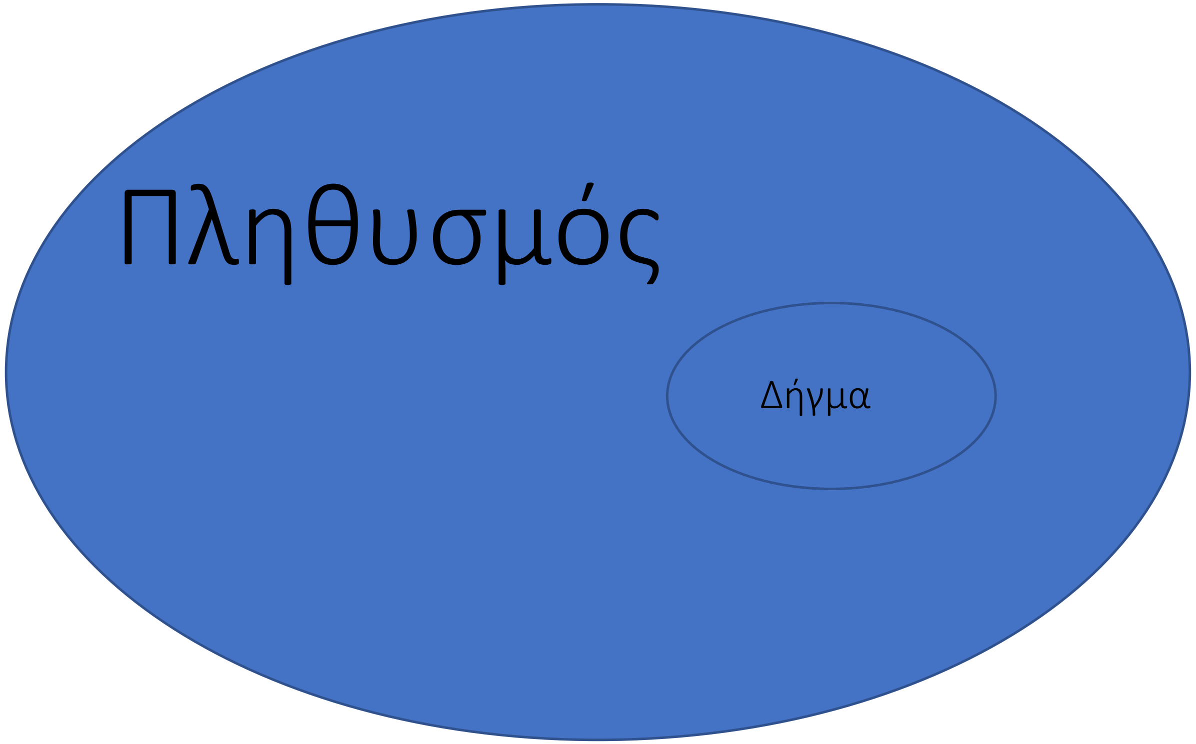
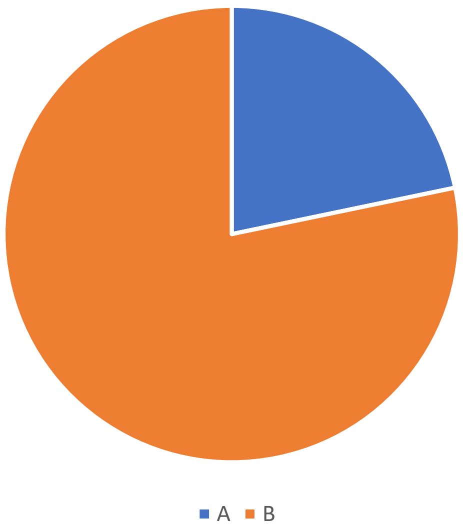
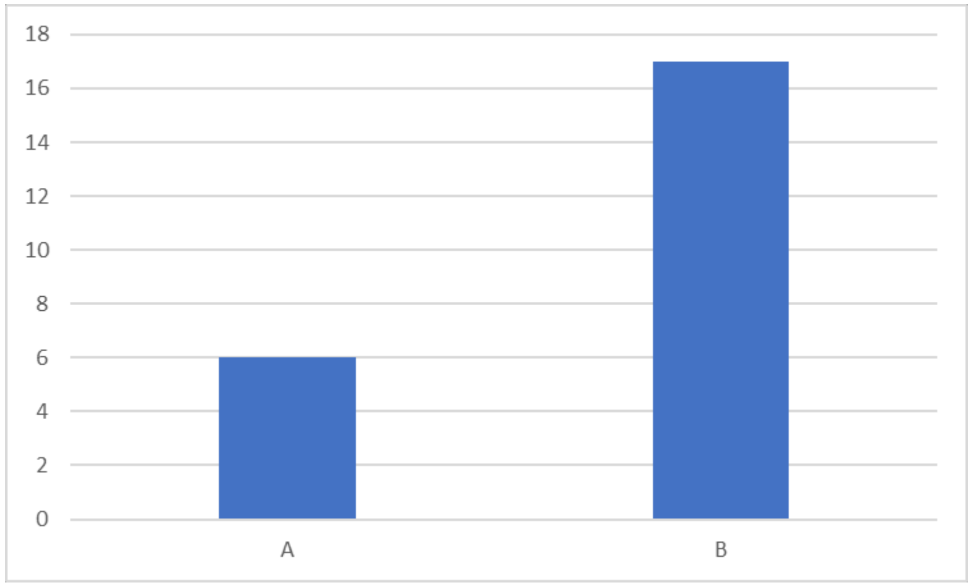
Στατιστική

Πληθυσμός - Δήγμα



Περιγραφική Στατιστική





Μέση τιμή και διακύμανση Δήγμα

Sample mean , Sample variance

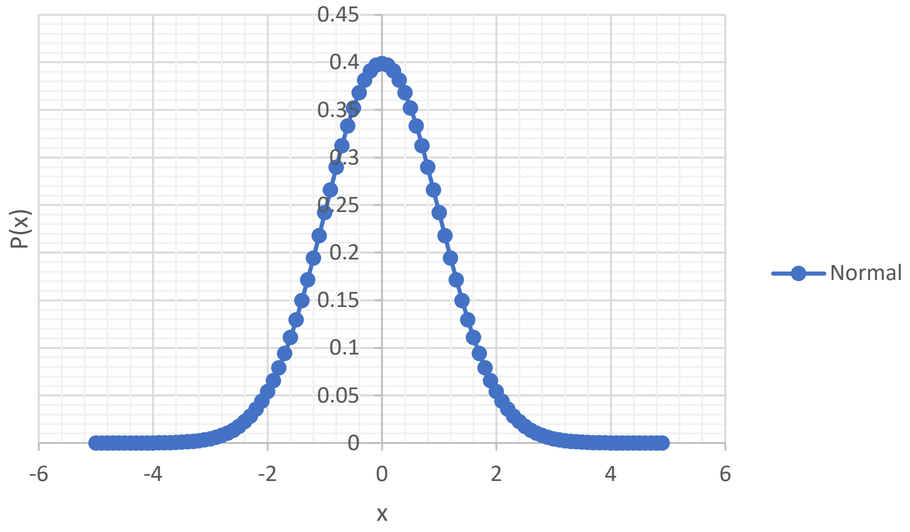
Δήγμα

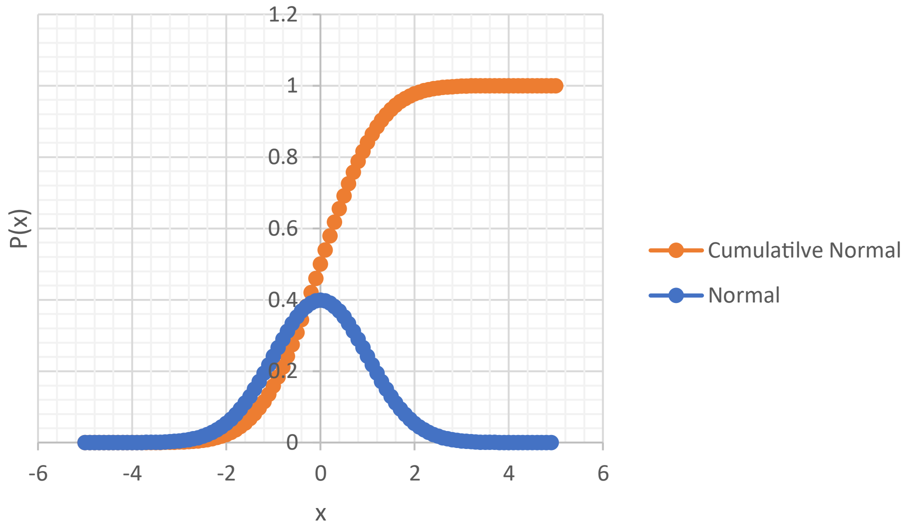
$$\langle x \rangle = \frac{\sum_{\iota}^{n} x_{\iota}}{n} \qquad s^2 = \frac{\sum_{\iota}^{n} \left(x_{\iota} - \langle x \rangle \right)^2}{n-1}$$

Πληθυσμός

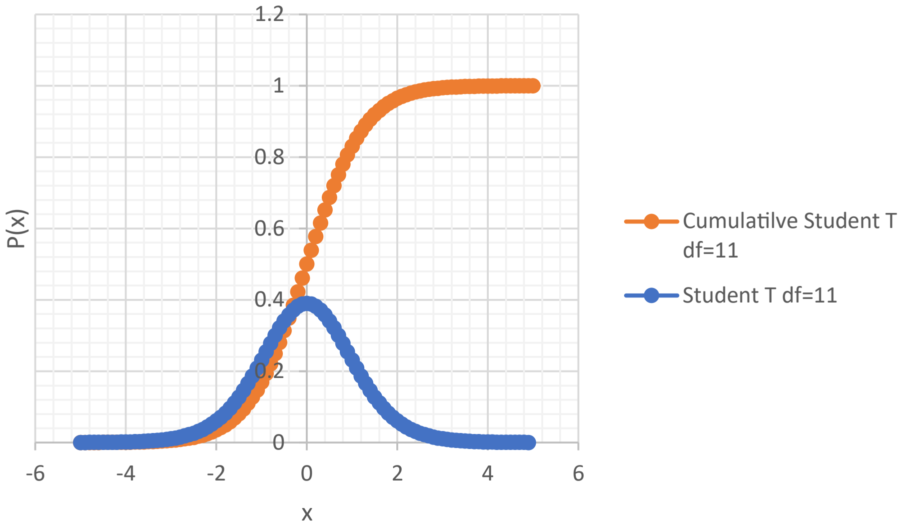
$$\mu = \frac{\sum_{\iota}^{n} x_{\iota}}{n} \qquad \sigma^2 = \frac{\sum_{\iota}^{n} \left(x_{\iota} - \langle x \rangle \right)^2}{n}$$

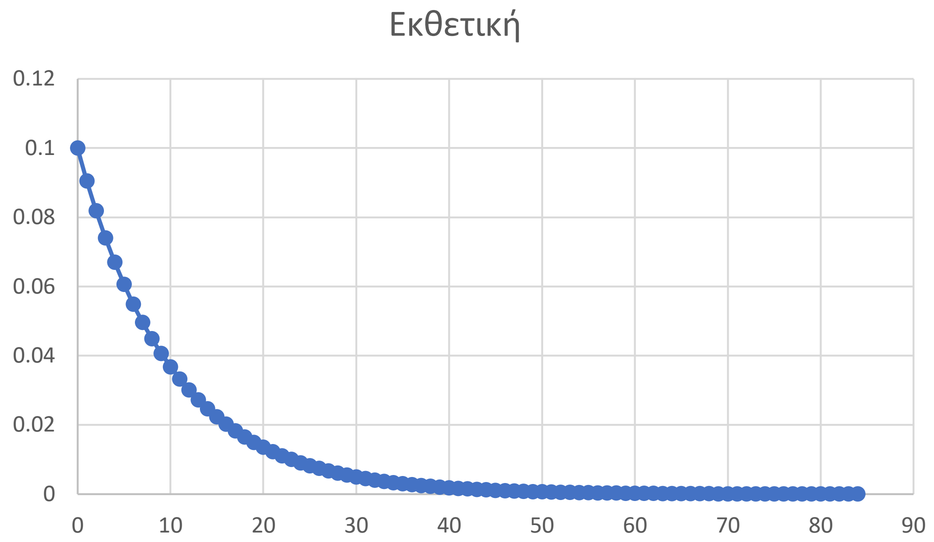
Κανονική κατανομή





Κατανομή Student T

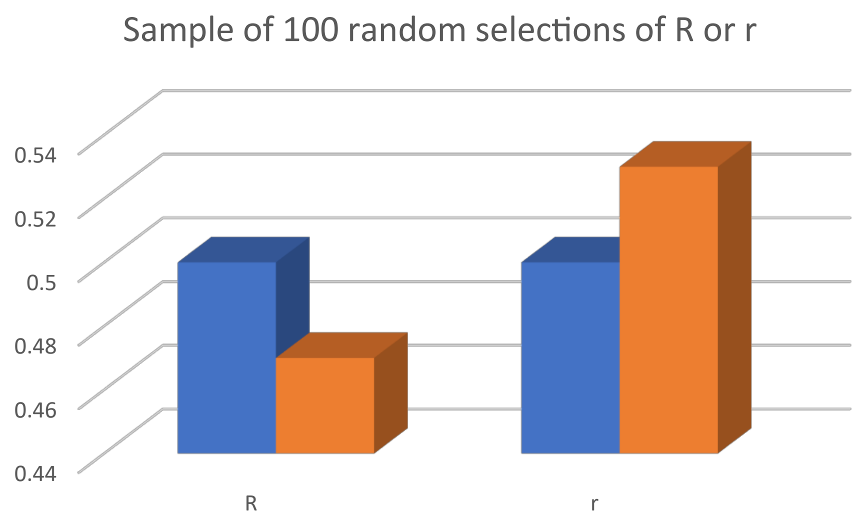


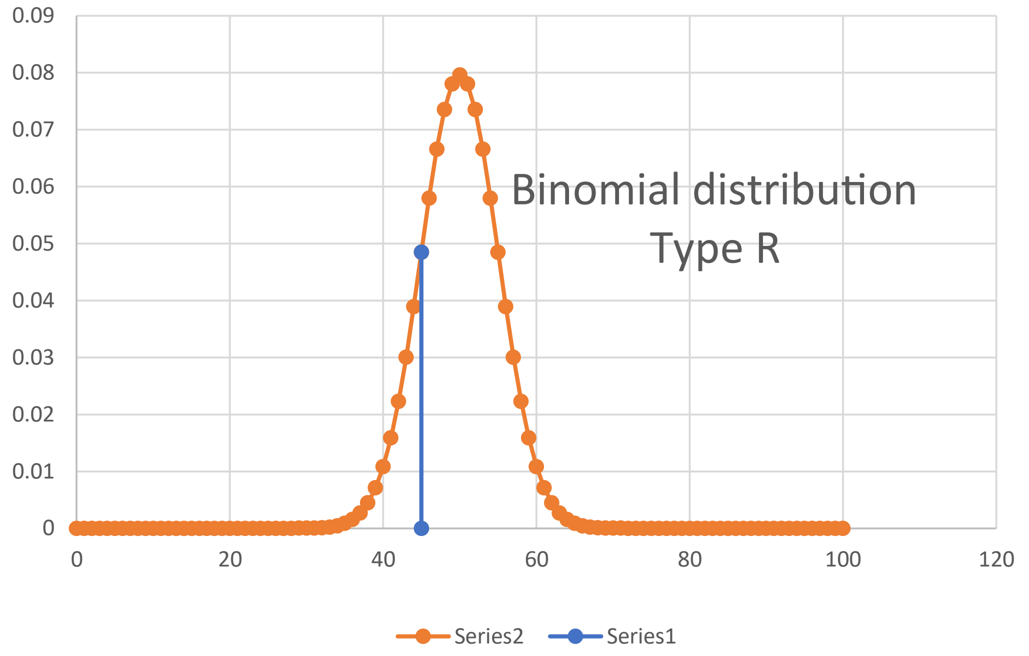

Εκθετική
0.12
0.1
0.08
0.06
0.04
0.02
0
0
10
20
30
40
50
60
70
80
90

Διωνυμική κατανομή

$$X_i = \text{ή} \begin{matrix} 0 \\ 1 \end{matrix}$$

$$P(\ X_i = Y) = \frac{N!}{Y! - (N - Y\,!)} p^Y (1-p)^{N-Y}$$



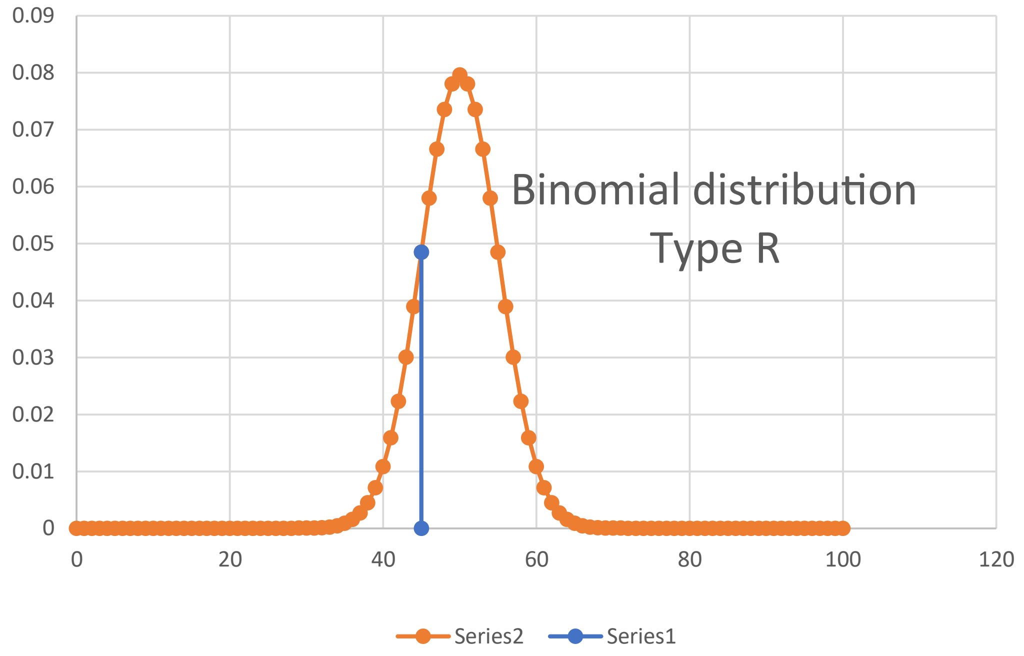


Διωνυμική κατανομή

$$\mu = < X \quad > = Np$$

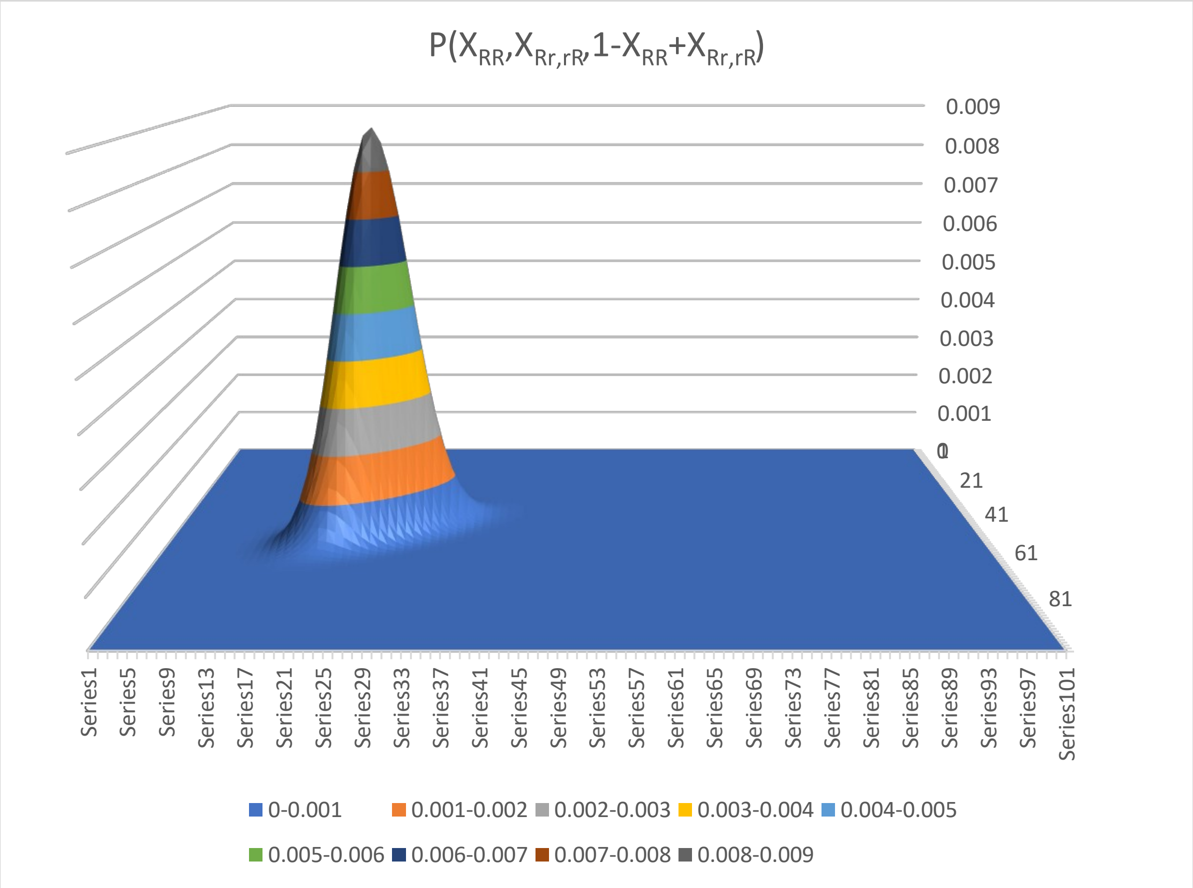
$$\sigma^2 = Np(1-p)$$

$$\sigma \quad = \sqrt{Np(1-p)}$$



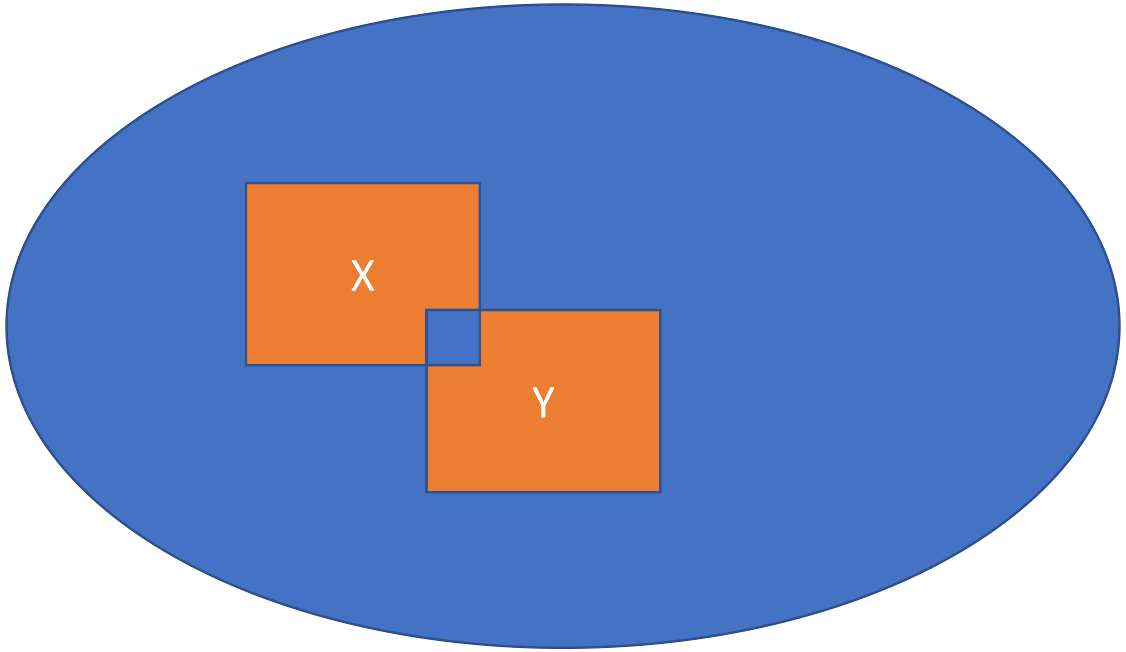
Multinomial Distribution

$$P(Y_1, .. Y_i, .. Y_k; p_1, p_i, p_k; N) = \frac{N!}{Y_1! \, Y_i! \, Y_k!} p_1^{Y_1} p_i^{Y_i} p_k^{Y_k}$$



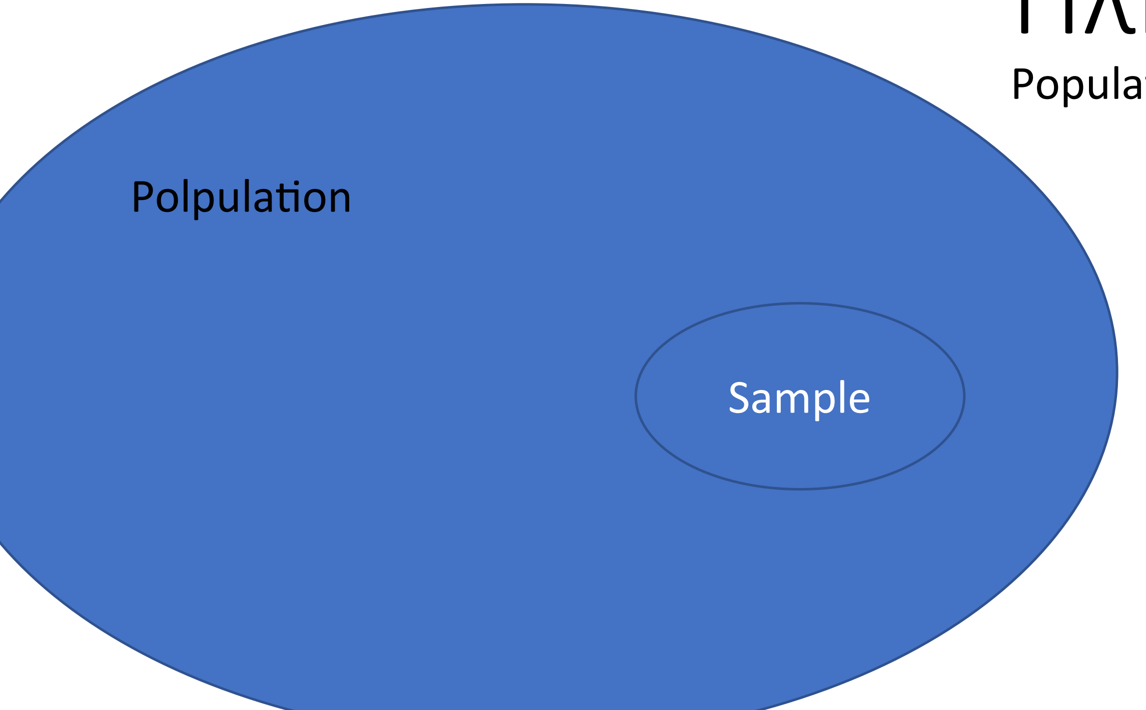
Υπο συνθήκη πιθανότητα

$$P\ (X \cap Y) = P\ (X)\ P(Y|X)$$



Πληθυσμός - Δήγμα

Sample vs Population



Κεντρικό οριακό θεώρημα

Center Limit theorem

Δήγμα

Sample

$$\langle x \rangle = \frac{\sum_i^n x_i}{n}$$

$$s^2 = \frac{\sum_i^n (x_i - \langle x \rangle)^2}{n-1}$$

Πληθυσμός

Population

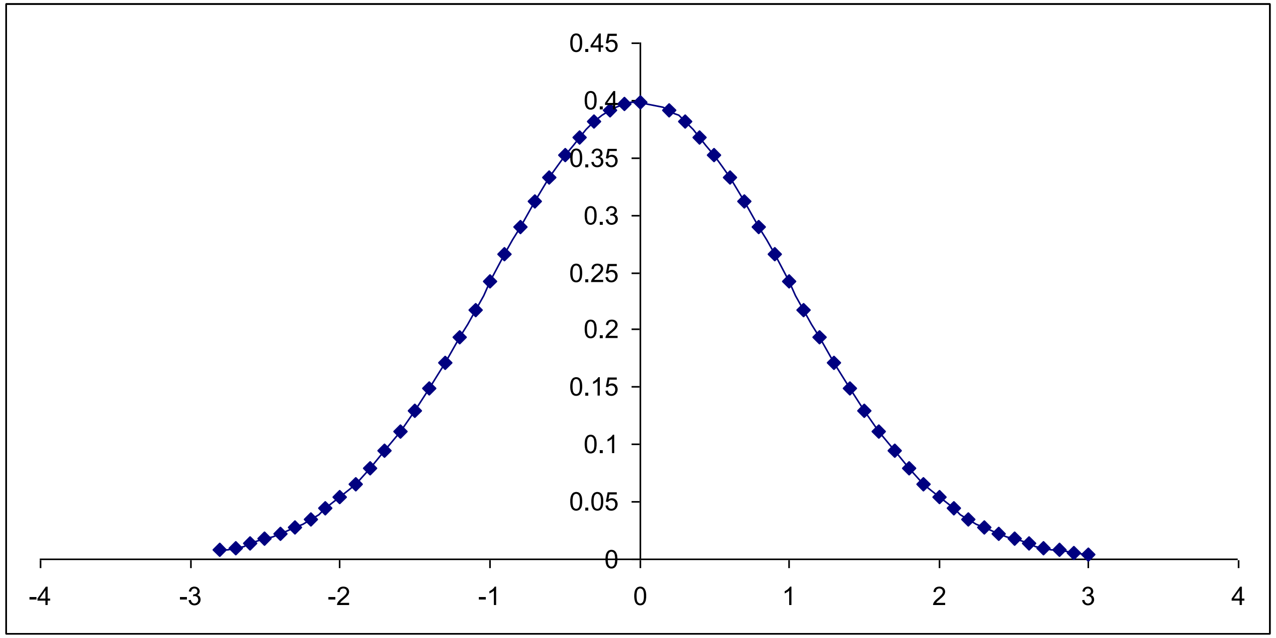
$$\mu = \frac{\sum_i^n x_i}{n}$$

$$\sigma^2 = \frac{\sum_i^n (x_i - \langle x \rangle)^2}{n}$$

$$\mu_{\langle x \rangle} = \mu$$

$$\sigma_{\langle x \rangle} = \frac{\sigma}{\sqrt{N}}$$

Κανονική κατανομή



$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Wikipedia

Τέστ υποθέσεων H_0 Και H_1

H_0 : **Μηδενική υπόθεση**
null hypothesis

H_1 : **Εναλλακτική υπόθεση**

alternative hypothesis

Επίπεδο σημαντικότητας
Significance **Level**

Wikipedia

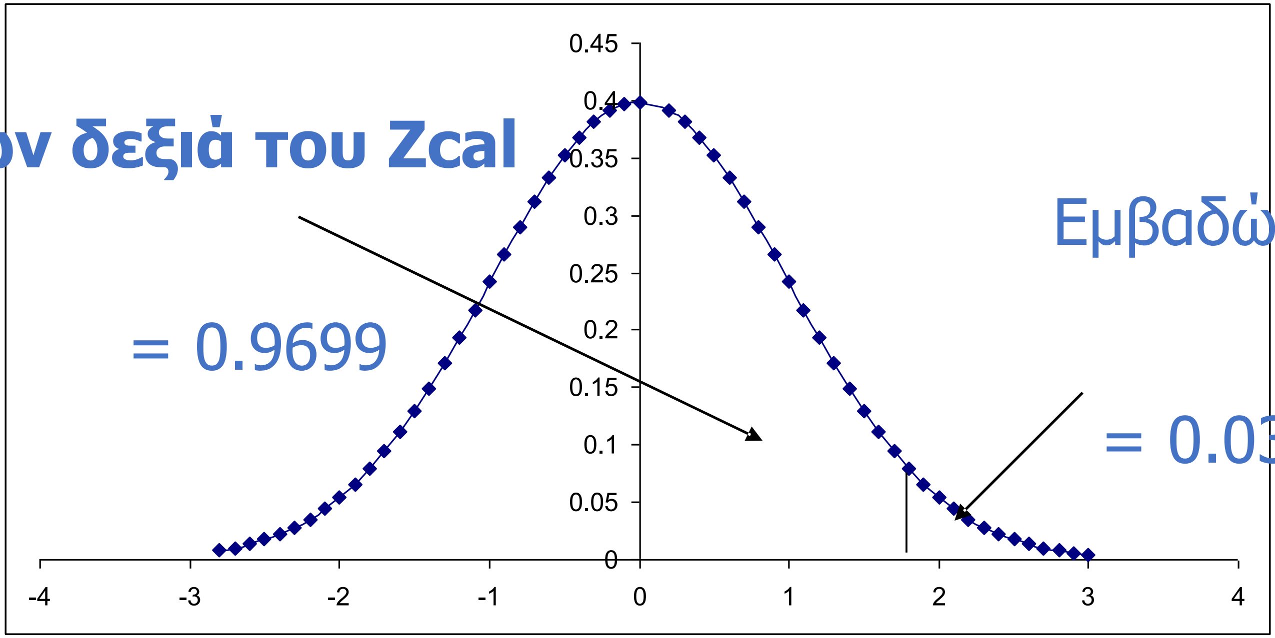
Balancing the two types of errors.

Κανονική κατανομή(Ζ τέστ)

Probability of z with a One-Tailed Test (Normal Distribution μ= 0, σ=1)

Εμβαδών δεξιά του Zcal

= 0.9699



Εμβαδών αριστερά του Zcal

= 0.0301 = p-value

$z_{CAL} = 1.88$

Determining the critical value of the test statistic, the area to the right of the critical value is either α or α/2. It is α for a one-tail test and α/2 for a two-tail test.

Κανονική κατανομή(Z τέστ)

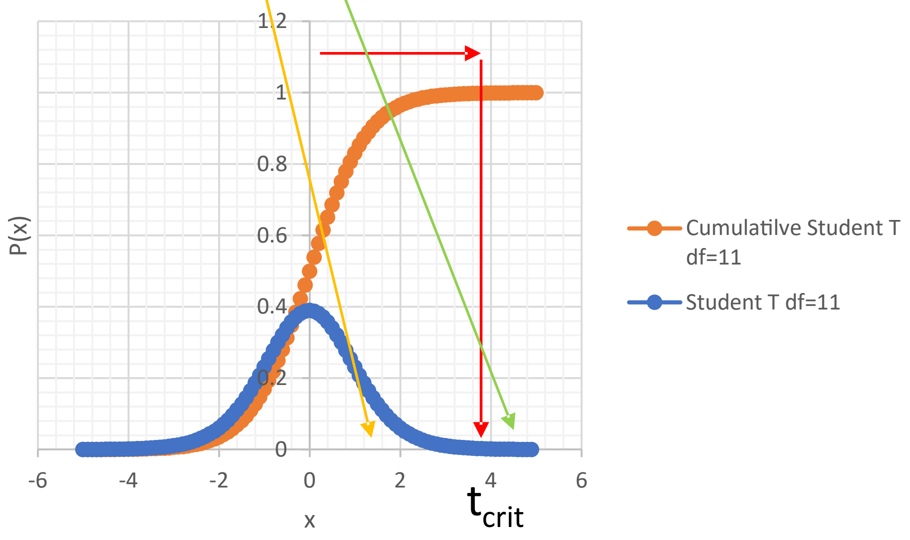
Using a t-Test on sample

$$t = \frac{(\langle x \rangle - \mu)}{\frac{s_X}{\sqrt{n}}}$$

df =n-1

accept

reject



Sample

$$\langle x \rangle = \frac{\sum_i^n x_i}{n} \qquad s^2 = \frac{\sum_i^n (x_i - \langle x \rangle)^2}{n-1}$$

Population

$$\mu = \frac{\sum_i^n x_i}{n} \qquad \sigma^2 = \frac{\sum_i^n (x_i - \langle x \rangle)^2}{n}$$

$$\mu_{\langle x \rangle} = \mu \qquad \sigma_{\langle x \rangle} = \frac{\sigma}{\sqrt{n}} \cong \frac{s}{\sqrt{n}}$$

Using a t-Test two samples (Independent)

Ίδιο μέγεθος δήγματος , αναμενόμενη ίση διακύμανση
equal size, expected equal σ

$$t = \frac{(\langle X_1 \rangle - \langle X_2 \rangle)}{s_p \sqrt{\frac{2}{n}}}$$

$$s_p = \sqrt{\frac{{s_{X_1}}^2 + {s_{X_2}}^2}{2}}$$

df=2n-2

Διαφορετικό μέγεθος δήγματος , αναμενόμενη ίση διακύμανση
Not equal size, expected equal σ

$$t = \frac{(\langle X_1 \rangle - \langle X_2 \rangle)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s_p = \sqrt{\frac{(n_1 - 1){s_{X_1}}^2 + (n_2 - 1){s_{X_2}}^2}{n_1 + n_2 - 2}}$$

df=2n-2

Διαφορετικό μέγεθος δήγματος , αναμενόμενη διαφορετική διακύμανση
Not equal size, expected NOT equal σ

$$t = \frac{(\langle X_1 \rangle - \langle X_2 \rangle)}{s_\Delta}$$

$$s_\Delta = \sqrt{\frac{{s_{X_1}}^2}{n_1} + \frac{{s_{X_2}}^2}{n_2}}$$

$$df = \frac{\left(\frac{{s_{X_1}}^2}{n_1} + \frac{{s_{X_2}}^2}{n_2}\right)^2}{\frac{\left(\frac{{s_{X_1}}^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{{s_{X_2}}^2}{n_2}\right)^2}{n_2 - 1}}$$

(t τέστ)

Using a t-Test two samples (dependent)

0 if we test if are μ_x=0

$$t = \frac{(\langle x \quad \rangle - \mu)}{\frac{s_X}{\sqrt{n}}}$$

$$\mathrm{x} = X_{\ 1} - X_2$$

df=n-1

Paired Samples in the two groups

Κεντρικό οριακό θεώρημα

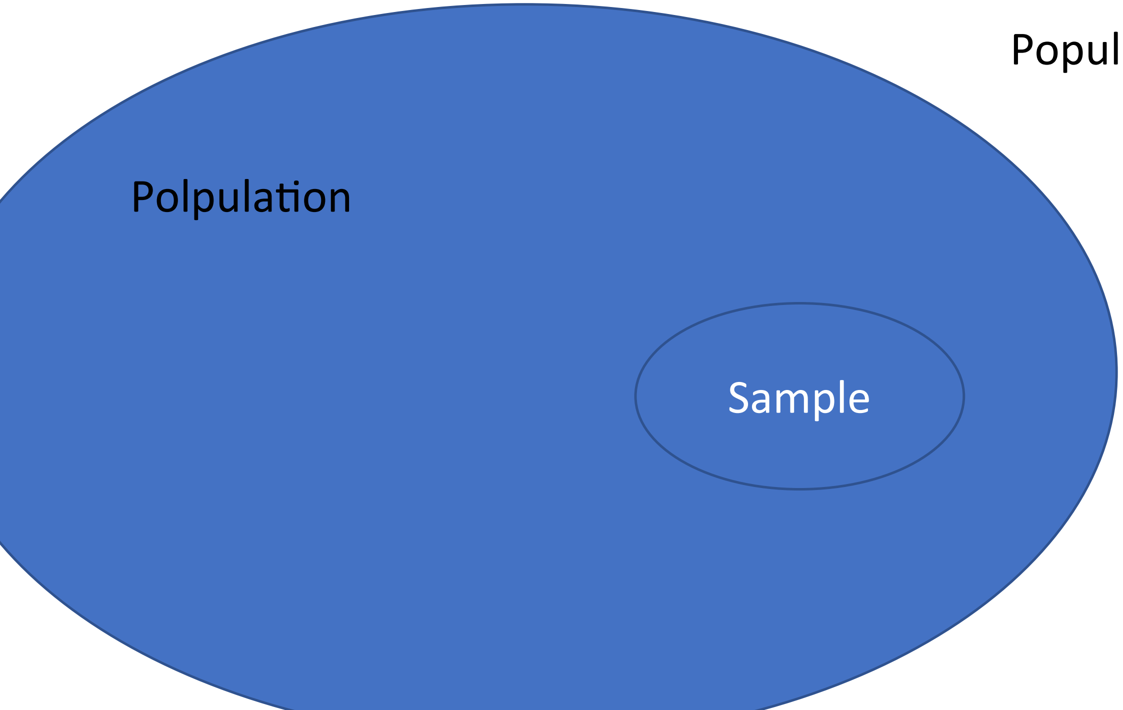
Center Limit theorem

Sample

$$\langle x \rangle = \frac{\sigma_{in}\, x_\iota}{n} \qquad s^2 = \frac{\sigma_i^n (x_\iota - \langle x \rangle)^2}{n-1}$$

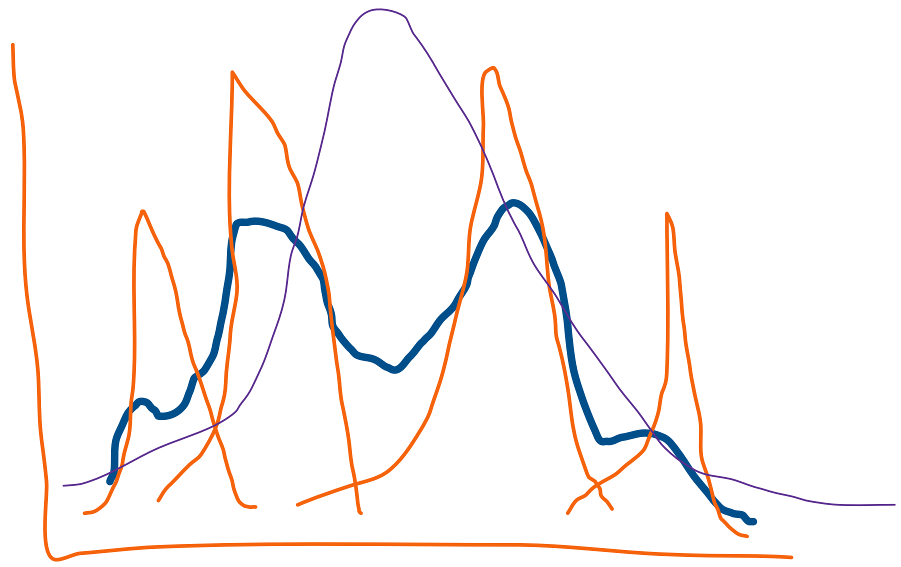
Population

$$\mu = \frac{\sigma_i^N\, x}{N} \qquad \sigma^2 = \frac{\sigma_i^N (x_\iota - \langle x \rangle)^2}{N}$$



$$\mu_{\langle x \rangle} = \mu \qquad \sigma_{\langle x \rangle} = \frac{\sigma}{\sqrt{N}}$$

Anova



Nonparametric test

Γενετική ισορροπία

Hardy-Weinberg Disequilibrium

$P_{aa} \neq p_a^2$ $\quad P_{Aa} \neq 2p_A p_a$ $\quad p_A + p_a = 1$

$D_{aa} = P_{aa} - p_a^2$ $\quad D_{Aa} = P_{Aa} - 2p_A p_a$

Is their a Correlation in Binomial distribution ?

$P_{aa} = p_a^2 + p_a(p_A)f$ $\quad P_{Aa} = 2p_A p_a(1-f)$

Let Xj , j = 1, 2 indicating whether the jth allele of a random individual is allele A or a.
Clearly, E(Xj) = p_a

$$Var(X_j) = p_i (1 - p_i)$$

$$Cov (X_i , X_j) = E (X_iX_j) - E (X_i) E (X_j) = P_{11} - p_1^2$$

$$Corr(X_i , X_j) = Cov(X_i , X_j)/(\text{sqrt } (\text{Var}(X_i)\ \text{Var}(X_j j))) = f$$

See examples in http://thirteen-01.stat.iastate.edu/wiki//stat536/files?filename=2008-09-09h.pdf

Γενετική ισορροπία

Hardy-Weinberg Disequilibrium

$$D_{aa} = P_{aa} - p_a^2 \qquad D_{Aa} = P_{Aa} - 2p_A p_a$$

But D's are not independent :

$$p_a = P_{aa} + \frac{1}{2} P_{Aa} \xrightarrow[P_{Aa} = D_{Aa} + 2p_A p_a]{} p_a = P_{aa} + \frac{1}{2}\left(D_{Aa} + 2p_A p_a\right)$$

$$\xrightarrow[D_{aa} + p_a^2 = P_{aa}]{} p_a = D_{aa} + p_a^2 + \frac{1}{2}\left(D_{Aa} + 2p_A p_a\right)$$

$$p_a = D_{aa} + p_a^2 + \frac{D_{Aa}}{2} + p_A p_a$$

$$p_a = D_{aa} + \frac{D_{Aa}}{2} + p_a\left(p_a + p_A\right) \longrightarrow D_{aa} = -\frac{D_{Aa}}{2}$$

Γενετική ισορροπία

Hardy-Weinberg Disequilibrium

$$P_{Aa} = D_{Aa} + 2p_A p_a$$

$$P_{aa} = D_{aa} + p_a^2$$

$$P_{AA} = D_{AA} + p_A^2$$

$$D_{AA} = D_{aa} = -\frac{D_{Aa}}{2}$$

$$\max(-p_a^2,\ -p_A^2) \le D_{AA} = D_{aa} \le p_A p_a$$

Testing $D_{AA} = 0$ (HWE)

Γενετική ισορροπία

Hardy-Weinberg Disequilibrium

Testing D_{AA} = 0 (HWE)

- An estimate D_{AA}
- A sampling distribution for the estimate D_{AA}.

$$N_{AA} = N(P_{AA}) = N(p_A^2 + D_{AA})$$

$$P_{aa} = D_{aa} + p_a^2$$

$$N_{Aa} = N(P_{Aa}) = N(2p_A p_a - 2D_{AA})$$

$$P_{Aa} = D_{Aa} + 2p_A p_a$$

Bailey’s method:

$$\hat{p}_A = \frac{2N_{AA} + N_{Aa}}{2N}$$

$$\hat{D}_{AA} = \frac{N_{AA}}{N} - \hat{p}_A^2 = \hat{P}_{AA} - \hat{p}_A^2$$

Γενετική ισορροπία

Hardy-Weinberg Disequilibrium

Testing $D_{AA} = 0$ (HWE)

Approximating

$$E\left[\hat{D}_{AA}\right] \approx 0$$

$$\text{var}\left[\hat{D}_{AA}\right] \approx \frac{1}{N}\left(\hat{p}_a^2 \hat{p}_A^2\right)$$

$$z = \frac{\hat{D}_{AA} - E\left[\hat{D}_{AA}\right]}{\sqrt{\text{var}\left[\hat{D}_{AA}\right]}} \approx \frac{N\hat{D}_{AA}}{\left(\hat{p}_a \hat{p}_A\right)}$$

Γενετική ισορροπία

Hardy-Weinberg Disequilibrium

Likelihood for HWE

$$\hat{p}_{AA} = \frac{N_{AA}}{N} \qquad \hat{p}_{Aa} = \frac{N_{Aa}}{N}$$

$$L = \left(\frac{N!}{N_{AA}!N_{Aa}!N_{aa}} \right) \left(\widetilde{P}_{Aa}\right)^{N_{Aa}} \left(\widetilde{P}_{AA}\right)^{N_{AA}} \left(\widetilde{P}_{aa}\right)^{N_{aa}}$$

Σχέση φαινοτύπου γονότυπου

Interface : form phenotype to genotype

Estimate allele frequencies in the ABO blood group system in humans

https://en.wikipedia.org/wiki/ABO_(gene)

Phenotype	**A**	**AB**	**B**	**O**
Genotype(s)	aa ao	ab	bb bo	oo
Sample	N_A	N_{AB}	N_B	N_O

Lecture Notes in Population Genetics : http://darwin.eeb.uconn.edu/eeb348/lecture-notes/book.pdf

Kent E. Holsinger Department of Ecology & Evolutionary Biology, U-3043 University of Connecticut Storrs, CT 06269-3043

Σχέση φαινοτύπου γονότυπο

Estimate allele frequencies in the ABO blood group system in humans

Phenotype	A	AB	B	O
Genotype(s)	aa ao	ab	bb bo	oo
Sample	N_A	N_{AB}	N_B	N_O

If the frequencies for the genotype where known :

$$N_{aa} = N_A \left(\frac{p_a^2}{p_a^2 + 2p_a p_o} \right) \qquad N_{ao} = N_A \left(\frac{2p_a p_o}{p_a^2 + 2p_a p_o} \right) \qquad N_{ab} = N_{AB}$$

$$N_{bb} = N_B \left(\frac{p_b^2}{p_b^2 + 2p_b p_o} \right) \qquad N_{bo} = N_B \left(\frac{2p_b p_o}{p_b^2 + 2p_b p_o} \right) \qquad N_o = N_o$$

Lecture Notes in Population Genetics : http://darwin.eeb.uconn.edu/eeb348/lecture-notes/book.pdf

Interface : form phenotype to genotype

Estimate allele frequencies in the ABO blood group system in humans

Phenotype	A	AB	B	O
Genotype(s)	aa ao	ab	bb bo	oo
Sample	N_A	N_{AB}	N_B	N_O

$$N_{aa} = N_A\left(\frac{p_a^2}{p_a^2 + 2p_a p_o}\right) \qquad N_{ao} = N_A\left(\frac{2p_a p_o}{p_a^2 + 2p_a p_o}\right) \qquad N_{ab} = N_{AB}$$

$$N_{bb} = N_B\left(\frac{p_b^2}{p_b^2 + 2p_b p_o}\right) \qquad N_{bo} = N_B\left(\frac{2p_b p_o}{p_b^2 + 2p_b p_o}\right) \qquad N_o = N_o$$

$$p_a = \left(\frac{2N_{aa} + N_{ao} + N_{ab}}{2N}\right)$$

$$p_b = \left(\frac{2N_{bb} + N_{bo} + N_{ab}}{2N}\right)$$

$$p_o = \left(\frac{N_{ao} + N_{bo} + 2N_{oo}}{2N}\right)$$

Lecture Notes in Population Genetics : http://darwin.eeb.uconn.edu/eeb348/lecture-notes/book.pdf

Phenotype	**A**	**AB**	**B**	**O**
Genotype(s)	aa ao	ab	bb bo	oo
Sample	NA	NAB	NB	NO
	12	44	33	22

Initial Estimates | | | | | | | | | Resulting

pa	pb	po	Naa	Nao	Nbb	Nbo	Noo	Nab	pa	pb	po
0.333333	0.333333	0.333333	4	8	11	22	22	44	0.27027	0.396396	0.333333
0.27027	0.396396	0.333333	3.461538	8.538462	12.30508	20.69492	22	44	0.267845	0.402275	0.32988
0.267845	0.402275	0.32988	3.464985	8.535015	12.49966	20.50034	22	44	0.26786	0.403152	0.328988
0.26786	0.403152	0.328988	3.471805	8.528195	12.53761	20.46239	22	44	0.267891	0.403323	0.328786
0.267891	0.403323	0.328786	3.473601	8.526399	12.54567	20.45433	22	44	0.267899	0.403359	0.328742
0.267899	0.403359	0.328742	3.474009	8.525991	12.54742	20.45258	22	44	0.267901	0.403367	0.328732
0.267901	0.403367	0.328732	3.474099	8.525901	12.5478	20.4522	22	44	0.267901	0.403368	0.32873
0.267901	0.403368	0.32873	3.474118	8.525882	12.54789	20.45211	22	44	0.267901	0.403369	0.32873

Prevalence

The number (or percentage) of people with a specific condition

Point Prevalence

The fraction of people with a specific condition at a given time

https://www.youtube.com/watch?v=L1LXQ_ZpM88&index=1&list=PL8NkkvvWlbpQJsQogbwo_ylI5YlIGwkst&t=0s

Incidence

The Likelihood of developing the disease within a period of time

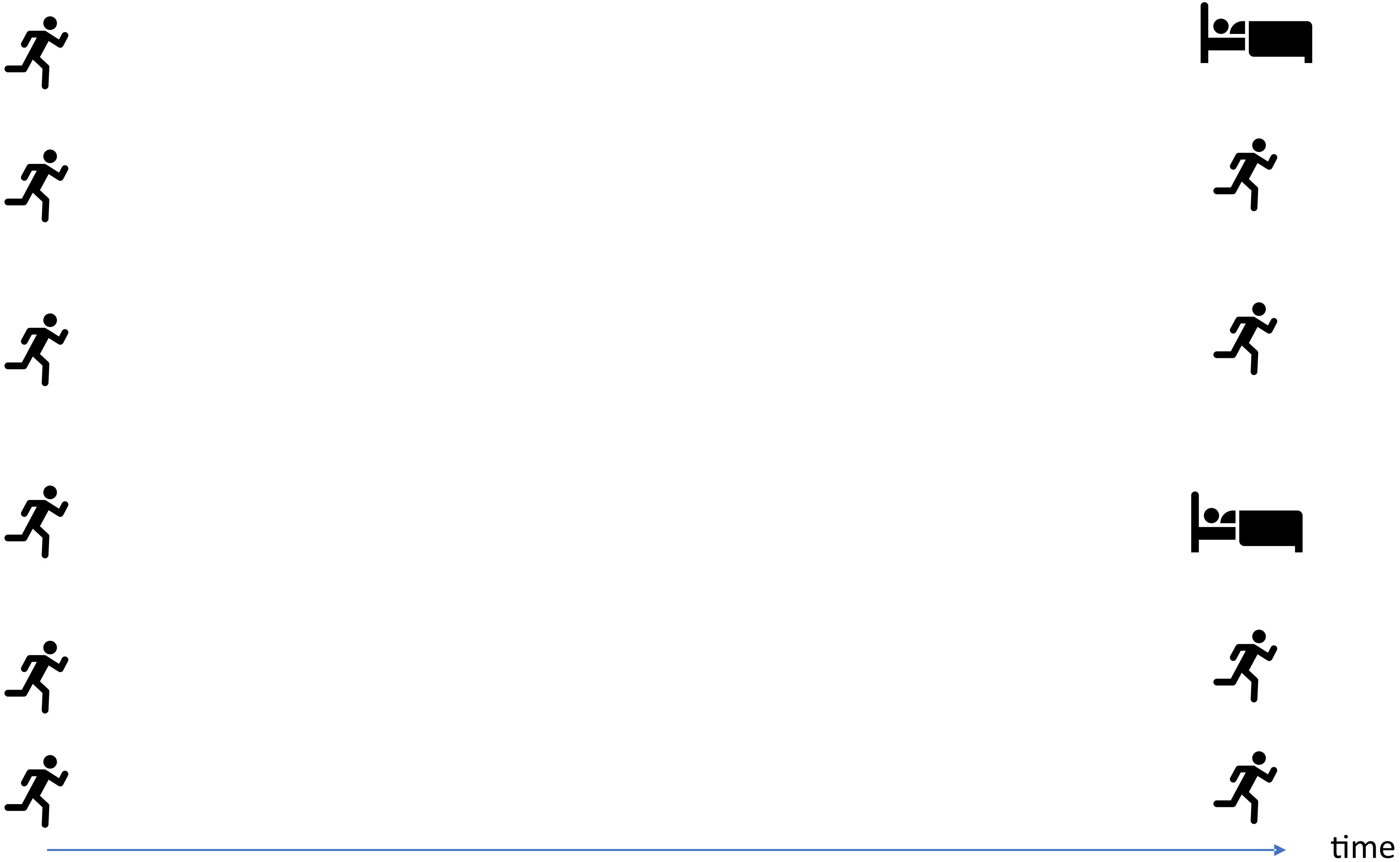
$$\text{Cumulative Incidence} : \frac{\text{The number of people the of developing the disease within a period of time}}{\text{The number of people at risk}}$$

$$\text{Incidence rate} : \frac{\text{The number of people the of developing the disease within a period of time}}{\text{The sum of length of time that the persons are free of disease}}$$

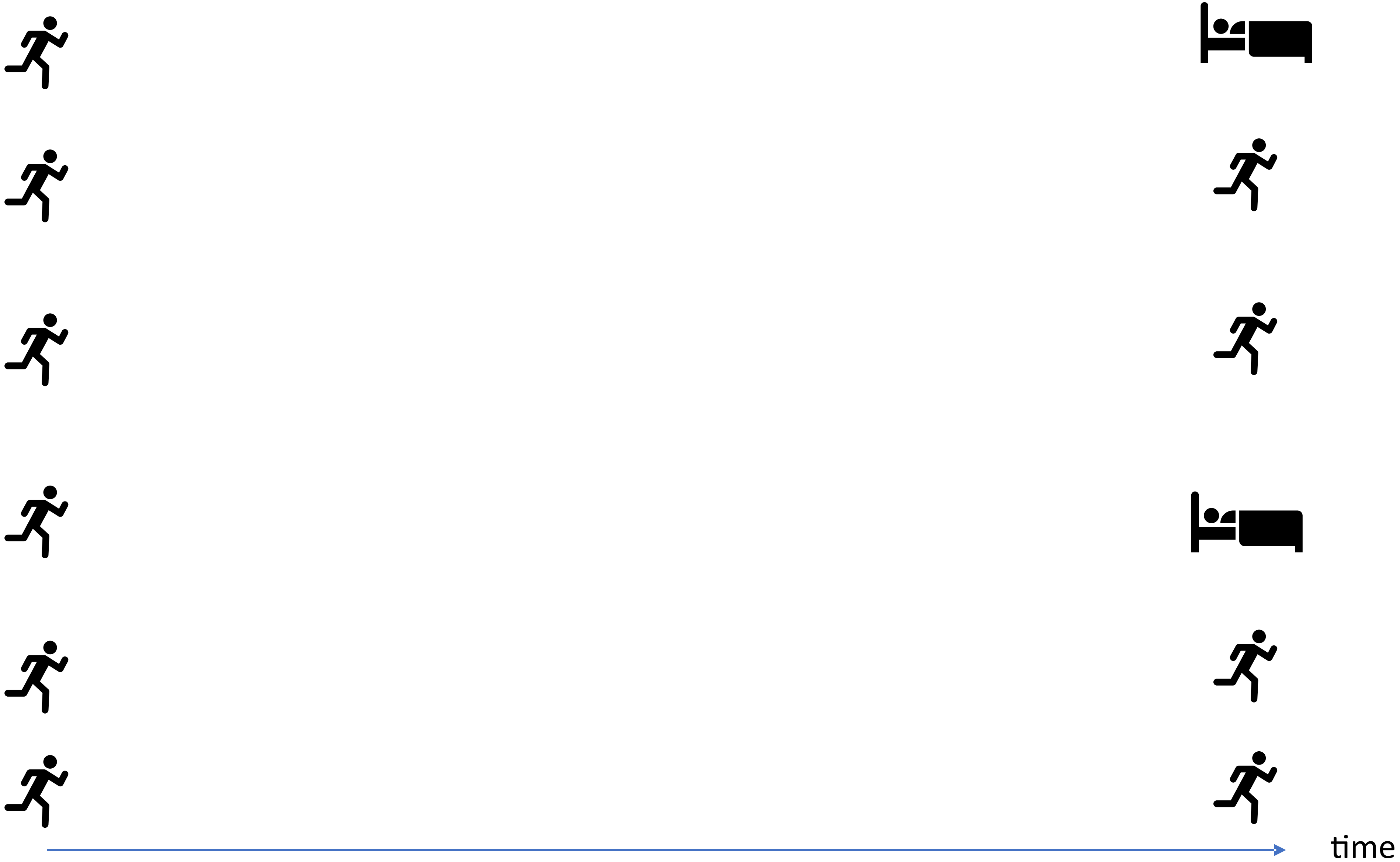
Cumulative Incidence : The number of people the of developing the disease within a period of time

The number of people at risk

=2/6



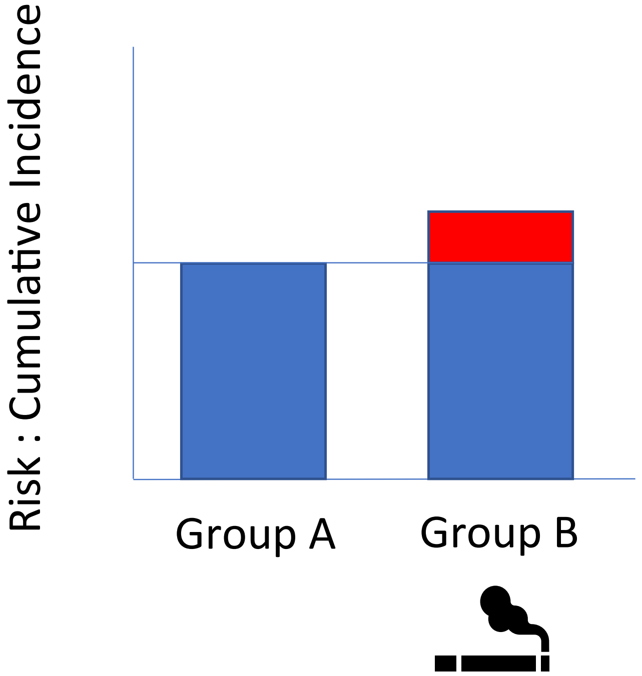
$$\text{Incidence rate : } \frac{\text{The number of people the of developing the disease within a period of time}}{\text{The sum of length of time that the persons are free of disease}} \quad =2/t$$



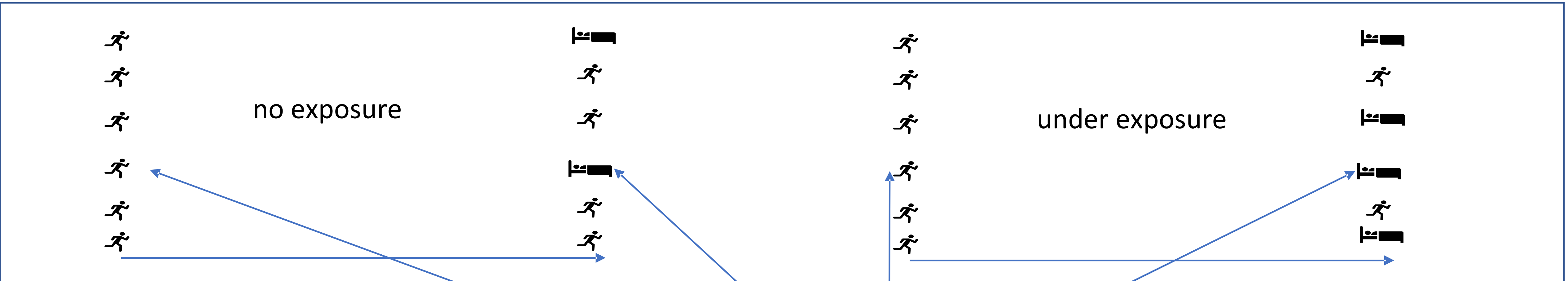
Compering Disease Risk between groups.

Risk deference
Relative Risk
Odds Ratio

Risk deference
In different Groups



Risk deference under exposure



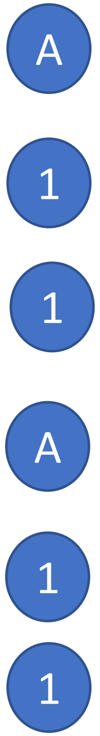
Attributable Risk = Incidence rate under expositor - Incidence rate with no exposure

	Not exposure	exposure	Total
Incidents	5	20	25
No Incidents	123	66	189
Total	128	86	214
% Incidents	4%	23%	12%
	Attribuble risk =	19%	
	Risk Fraction =	83%	
	Relative risk	5.95	

Probubility vs odds and odds Ratio

Probubility 2/6

Relative odds 2/4

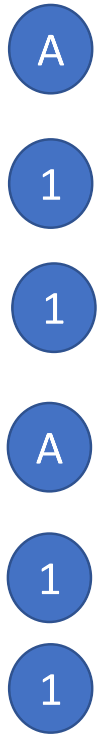


http://uncyclopedia.wikia.com/wiki/File:Flipping-coin-animated.gif

odds Ratio (exposed vs not exposed)

Not-exposed

2/4



exposed

2/3

odds Ratio =2/3/ (2/4) = 4/3

V

Relative risk

	Not exposure	exposure	Total
Incidents	5	20	25
No Incidents	123	66	189
Total	128	86	214
% Incidents	4%	23%	12%
	Attribuble risk =	19%	
	Risk Fraction =	83%	
	Relative risk	5.95	

Odds Ratio

	Not exposure	exposure	Total
Disease	5	20	25
No Disease	123	66	189
Total	128	86	214
odds for Disease	0.040650407	0.30303	
odds ratio			7.454545

	Not exposure	exposure	
Incidents	**5**	**20**	25
No Incidents	**123**	**66**	189
	128	86	
	3.90625	23.25581395	
		Attribuble risk	19.34956395
		Relative risk	5.953488372
		odd ratio	7.454545455

Ανεξάρτητες μεταβλητές

$$P(Y|X) = P(Y)$$

$$P(X \cap Y) = P(X)P(Y|X)$$

Sensitivity Specificity of a test

- Y : positive test
- X : have the disease

Sensitivity $P(Y = true | X = true)$ $= \frac{\#\ (Y = true) \cap (X = true)}{\#(X = true)}$

Specificity $P(Y = false | X = false)$

False negative fraction : $P(Y = false | X = true)$

False positive fraction : $P(Y = true | X = false)$

$P(X = true|Y = true)$

$X = true$ $X = false$

$Y = true$ $\# \, (Y = true \cap X = true)$ $\# \, (Y = true \cap X = false)$

$Y = false$ $\#(Y = false \cap X = true)$ $\#(Y = false \cap X = false)$

$P(Y = true|X = true)$

$P(Y = false|X = false)$

https://en.wikipedia.org/wiki/Sensitivity_and_specificity

Bayes Theorem

$$P(X \cap Y) = P(Y|X)P(X) = P(X|Y)P(Y)$$

$$P(Y|X) = \frac{P(Y)}{P(X)} P(X|Y)$$