Fundamentals of Biostatistics

Principle of variability
Probabilistic analysis approach
Inferences from sample data to a population

Pr Roch Giorgi

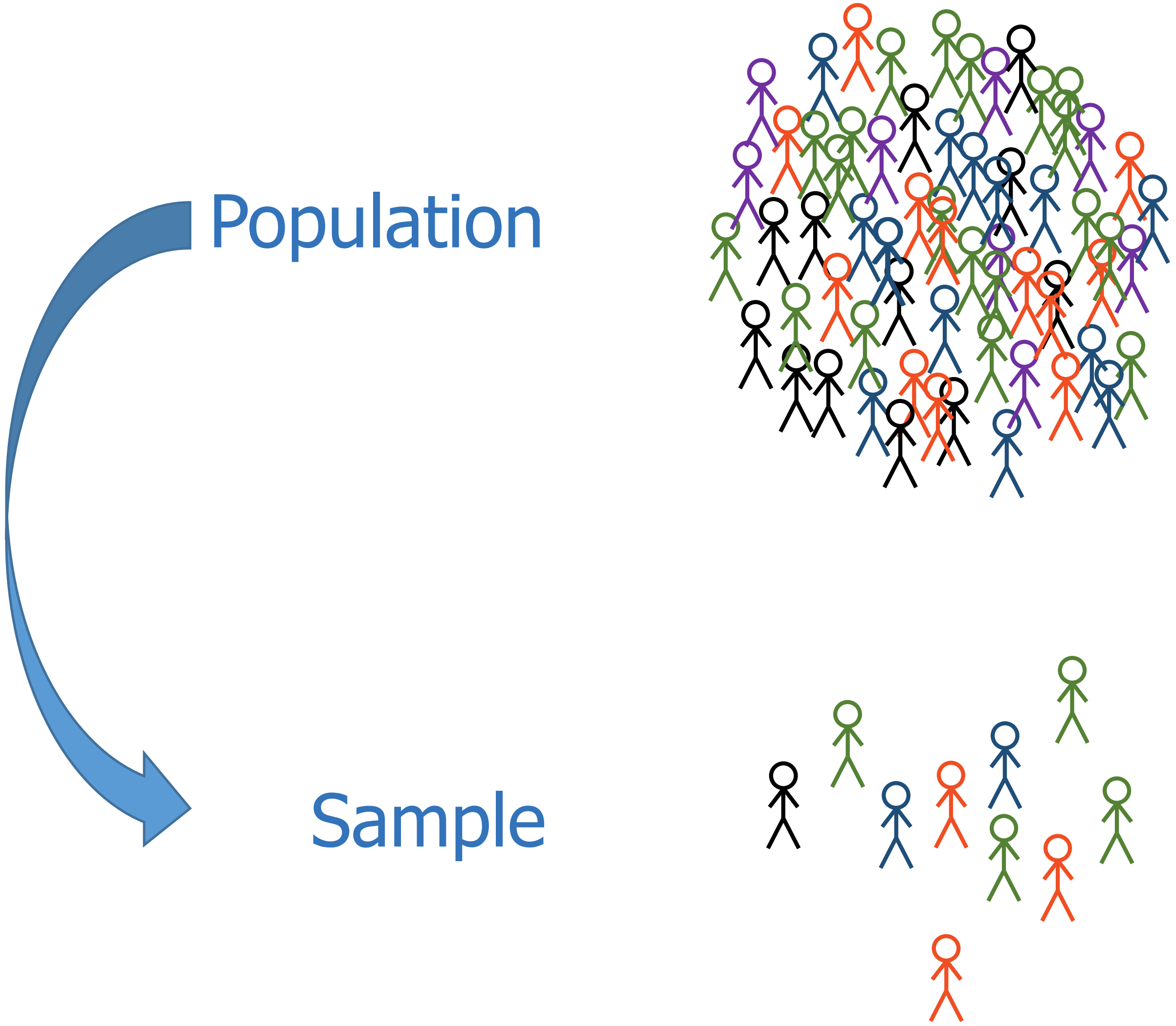
roch.giorgi@univ-amu.fr

SESSTIM, Faculty of Medical and Paramedical Sciences, Aix-Marseille University, Marseille, France
https://sesstim.univ-amu.fr/

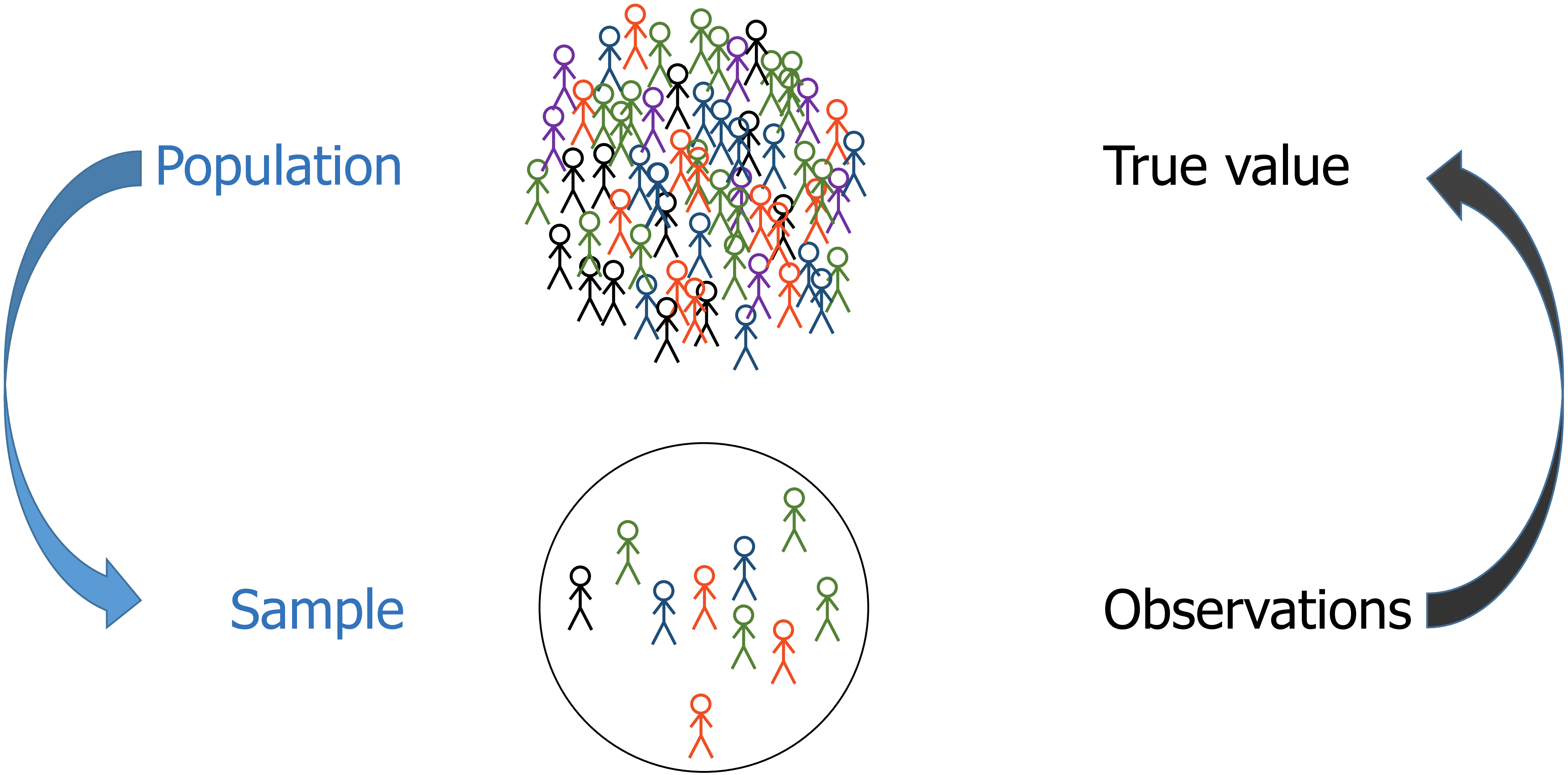
Population and Sample: Definitions

- Population
  - Set of individuals with their own characteristics
    - Individuals aged 75 and over with atrial fibrillation
  - Number of individuals often high
    - Prevalence around 400,000 to 700,000 in French
- Sample
  - Subset of a population
  - On each individual of the sample, one characteristic can be measured which is the subject of the study (often impossible on the whole population)
    - Occurrence of stroke, in order to estimate for each individual its own risk of stroke, or of other embolic event, the main predictors features (variables),…

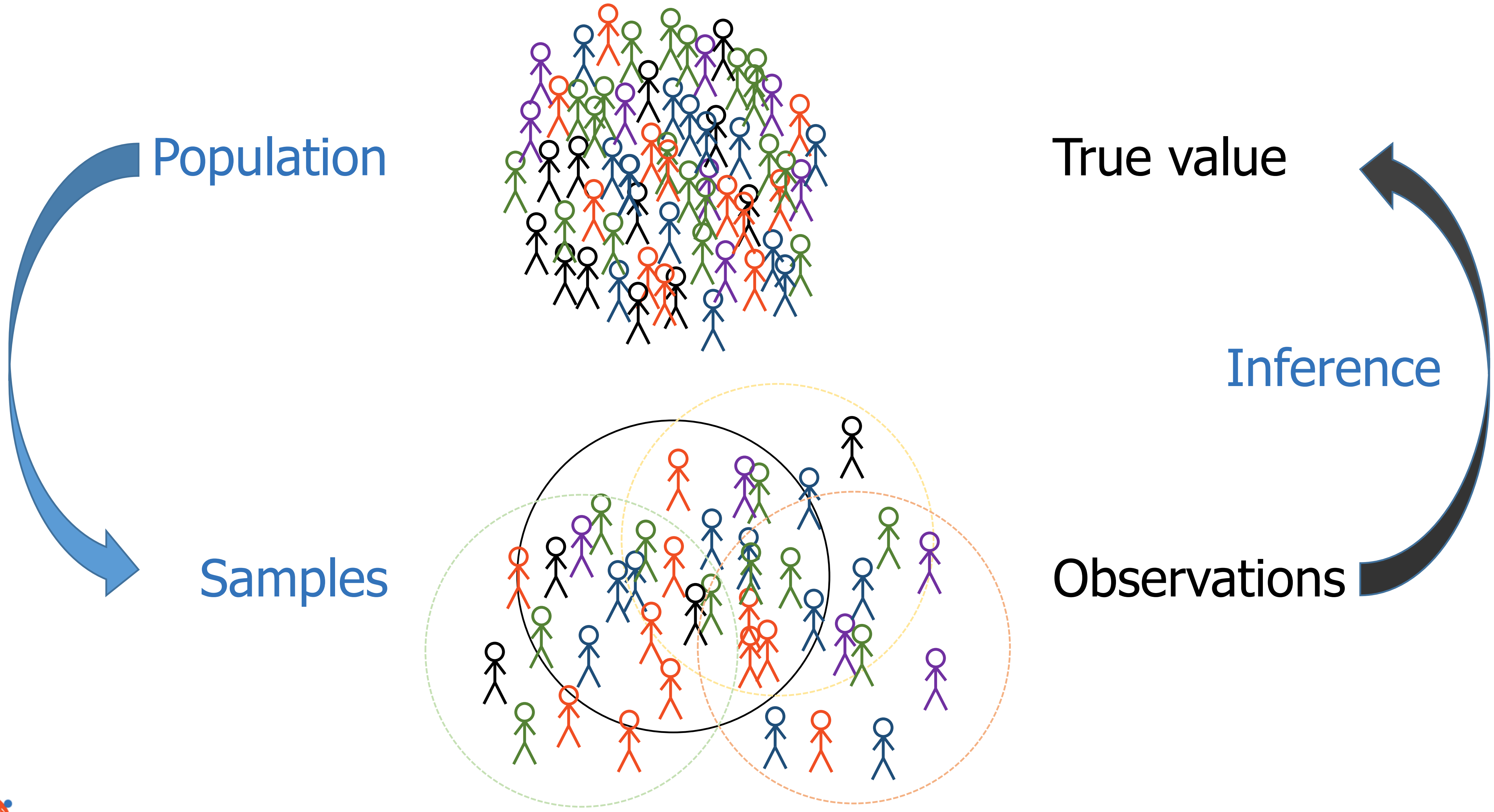
Population and Sample



Population and Sample



Population and Sample



Sample

- Observations made on the sample are used to answer questions about the target population
- Observed characteristics are random variables
- Their descriptive parameters allow us to know the distribution in the target population
    - Objective: estimate the parameters of the target population distribution
    - Way: use the observations made on the sample

Population and Sample

Population

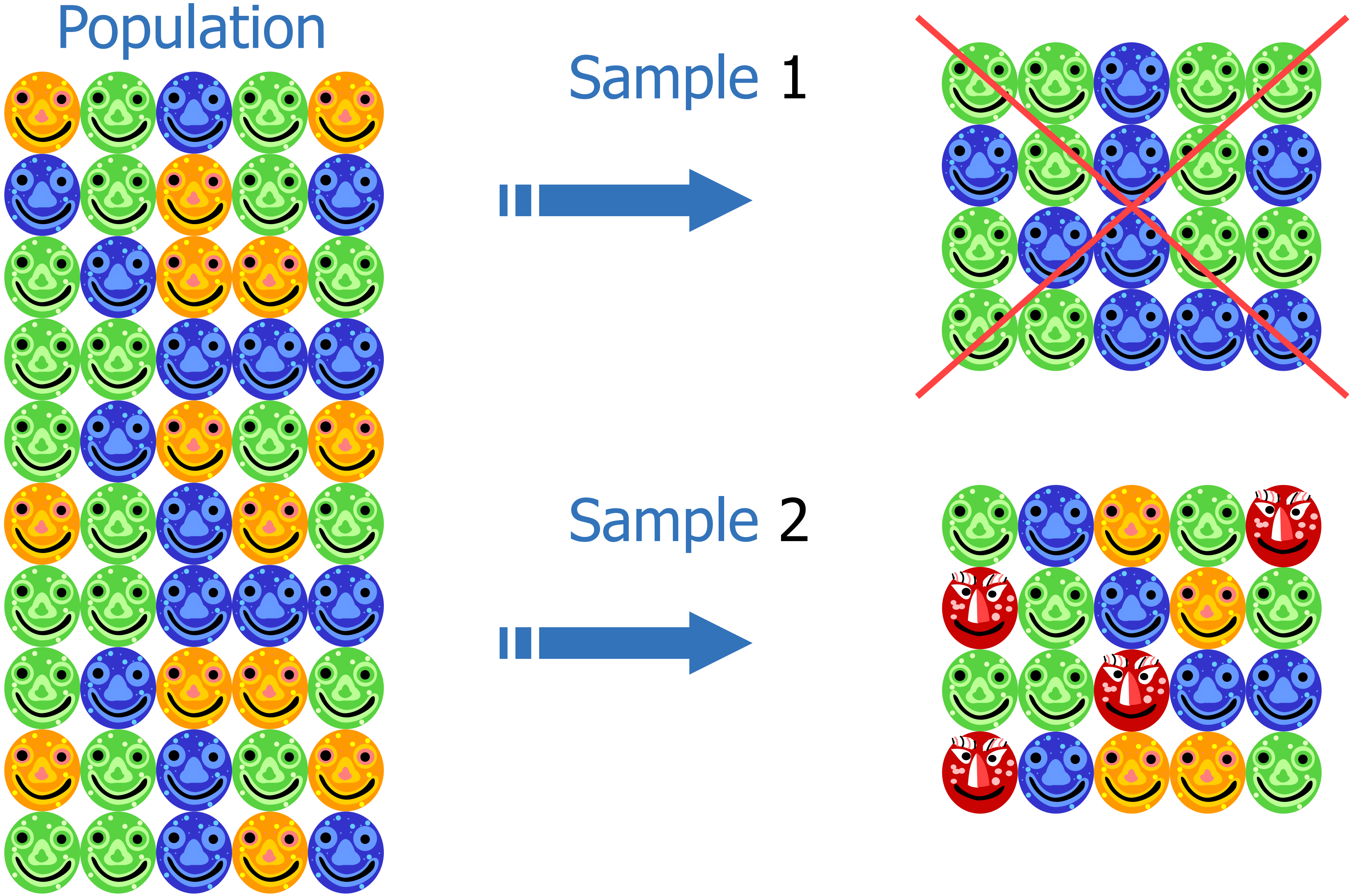
Sample 1

Population and Sample

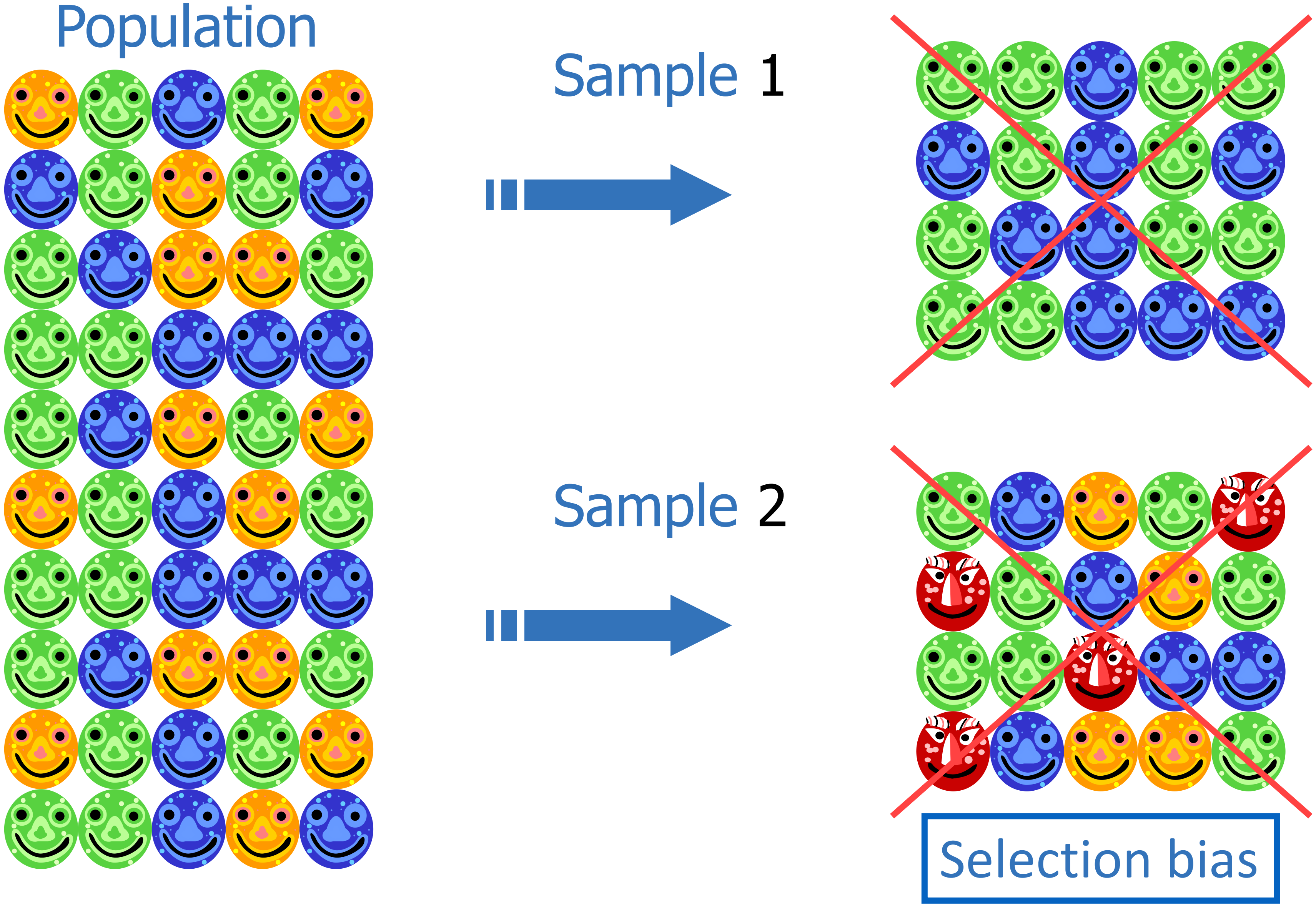
Population

Sample 1

Population and Sample



Population and Sample



Population and Sample

Population

Sample 3

= 28%

= 40%

= 32%

= 65%

= 15%

= 20%

Population and Sample

Population

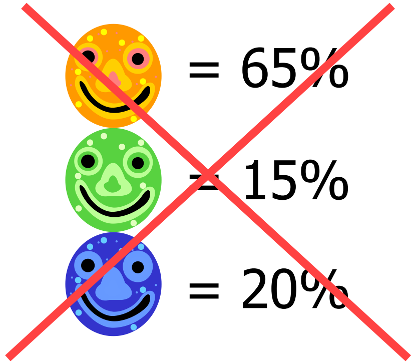
Sample 3

= 28%

= 40%

= 32%

≠



Selection bias

Population and Sample

Population

Sample 4

= 28%

= 40%

= 32%

≈

= 25%

= 40%

= 35%

Representative sample

Population and Sample

Sample		Population
Criterion of interest *Estimation of Feature A* - *Mean age, standard deviation* - *Probability of atrial fibrillation at 5 yr* - *...*	*Inference* →	Criterion of interest Feature A ? - Age - Atrial fibrillation at 5 yr - ...

Creating the Sample-s

- A sample provides information on the population
- A good ("unbiased") sample should be representative of the population from which it is drawn
- Need to precisely define the population
- Random sampling (pick at random) is the best way to do this
- The choice of the process may depend on the objective of the study, and therefore on the design of the study
  - Selection of case / of control in a case-control study
  - Selection of individuals in a retrospective cross sectional study
  - Selection, creation of patient groups in a prospective randomised controlled trial
  - Selection, creation of the exp. / non-exposed groups in a prospective cohort study

Sample and Representativeness

- Selection method defined *a priori*
- Description of the study subjects
- Selection criteria
  - Inclusion, non-inclusion criteria
- Deviations from the protocol

Random Sampling

- Each individual in the population has an equal chance of being in the sample (equiprobability)

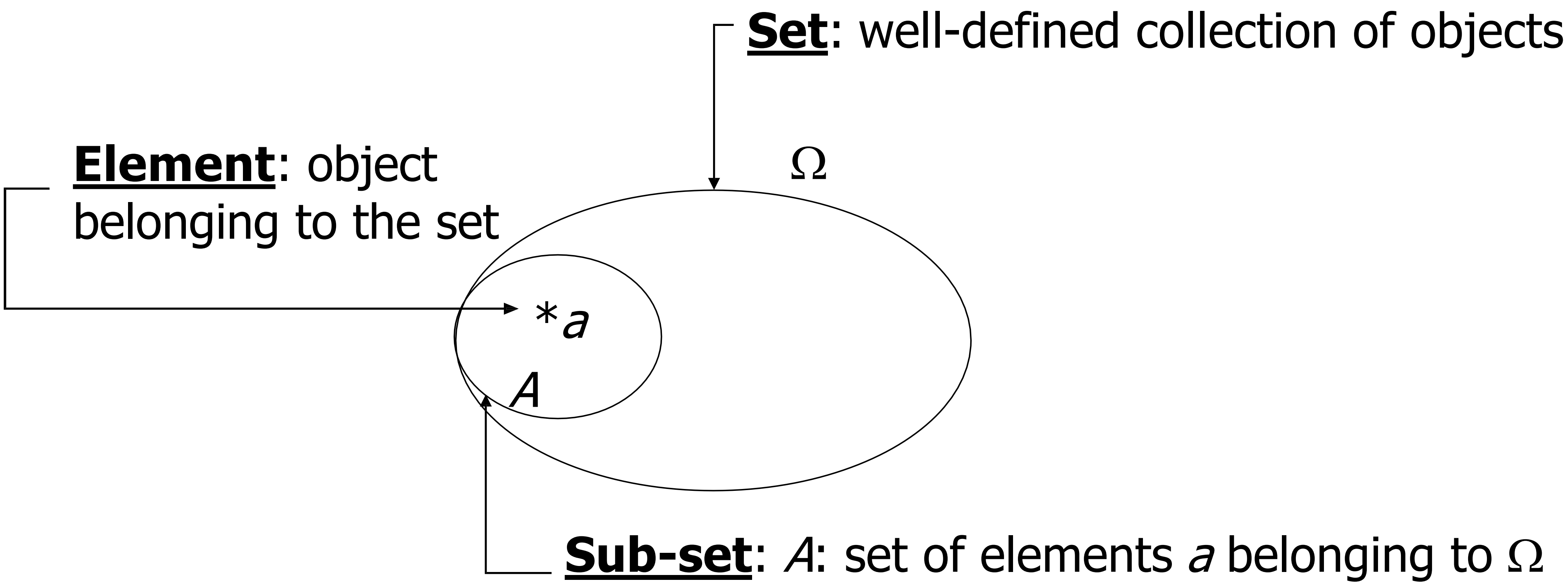
- Sampling
  - Simple
  - Stratified (ex.: centre, sex,…)

Probabilities

- Probability: models random phenomena whose outcomes are known but whose value cannot be predicted because their realisation is uncertain

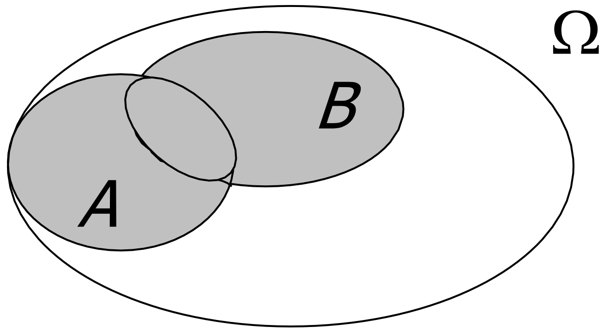
- Observation of the outcomes of a random phenomenon on sufficiently large series allows to determine their frequencies and subsequently the distribution that governs it

Reminders about Sets



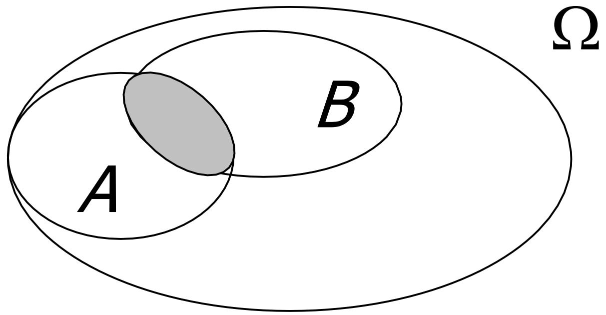
Reminders about Sets

<u>Union</u>: $A \cup B \Leftrightarrow A$ **or** B

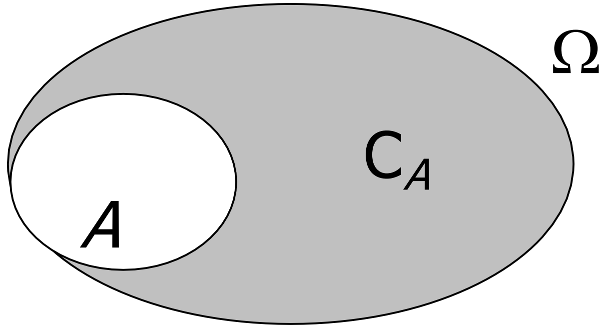


<u>Intersection</u>: $A \cap B \Leftrightarrow A$ **and** B

if $A \cap B = \varnothing$ then A and B are disjoints



<u>Complementarity</u>: C_A



Notion of Probability

- Probability: modelling of random phenomena

- Universal set, Ω: set of possible outcomes (all objects) for a given experiment (certain event)

- Event: subset *A* of Ω, that is a collection of outcomes (objects). An elementary event is *a*

Example
Rolling a 6-sided faire dice

Universal set, Ω = {f1, f2, f3, f4, f5, f6}

Event *A*: faces of number ≤ 2 = f1 $\cup$ f2

Event *B*: faces of number ≥ 5 = f5 $\cup$ f6

Event *C*: faces of even number {2, 4, 6} = f2 $\cup$ f4 $\cup$ f6

$A \cup B = f1 \cup f2 \cup f5 \cup f6, A \cap B = \varnothing$

$A \cup C = f1 \cup f2 \cup f4 \cup f6 , A \cap C \neq \varnothing$

Notion of Probability

Experiment repeated n times

	f1	f2	f3	f4	f5	f6	Total
Absolute Frequencies	n_1	n_2	n_3	n_4	n_5	n_6	n
Relative Frequencies (Fr.)	n_1/n	n_2/n	n_3/n	n_4/n	n_5/n	n_6/n	1

$Fr.(A) = (n_1+n_2)/n$

$Fr.(A \cup B) = (n_1+n_2+n_5+n_6)/n = (n_1+n_2)/n + (n_5+n_6)/n = Fr.(A) + Fr.(B)$

$Fr.(A \cup C)=(n_1+n_2+n_4+n_6)/n \neq Fr.(A) + Fr.(C)$

When $n \rightarrow \infty$ the relative frequency of an event tends towards the probability of that event

Probability Axioms

- Let be Ω a fundamental set, P the probability function that associates to any event A a positive or null real number. P(A) is called the probability of event A if:

$$P(A) \geq 0$$

$$P(\Omega) = 1$$

$$\text{if } A \cap B = \varnothing \Rightarrow P(A \cup B) = P(A) + P(B)$$

$$\text{if } A_i \cap A_j = \varnothing \Rightarrow P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$$

It can be deduced that:

$$P(\varnothing) = 0$$

$$P(A) \leq 1$$

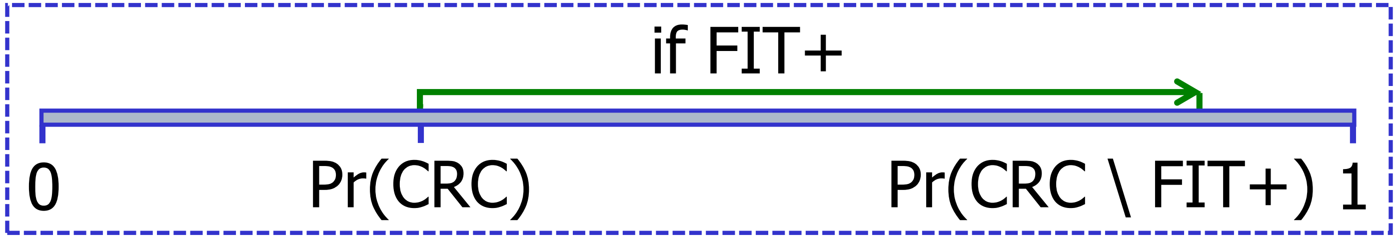
$$P(C_A) = 1 - P(A)$$

$$\text{if } A \subset B \text{, the } P(A) \leq P(B)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Conditional Probability

- Example
  - We are interested in the faecal immunochemical test (FIT) for colorectal cancer (CRC) screening
  - The probability of having CRC knowing that the FIT is positive is a conditional probability: P(CRC \ FIT+)



Conditional Probability

- The probability of *A* knowing *B* is defined by

$$P(A\backslash \mathrm{B}) = \frac{P(A \cap B)}{P(B)}$$

hence $P(A \cap B) = P(A\backslash \mathrm{B})P(B) = P(B\backslash \mathrm{A})P(A)$
and

$$P(A\backslash \mathrm{B}) = \frac{P(B\backslash \mathrm{A})P(A)}{P(B)}$$

Bayes' rule

Independence in Probability

- A and B are independents if and only if

$$P(A \cap \mathrm{B}) = P(A)P(B)$$

- If A and B are independents and $P(A) > 0$, $P(B) > 0$, then

$$P(A \backslash \mathrm{B}) = P(A \cap B)/P(B) = P(A)P(B)/P(B) = P(A)$$

- 2 disjoints events with non-null probabilities are never independent
  - Disjoints: $P(A \cap B) = 0$
  - Independents: $P(A \cap B) = P(A)P(B)$

Conditional Probability

- $A_1,\ldots,A_n$ events that partition Ω
- B any event

then

$$P(B) = P(B \cap A_1) \cup P(B \cap A_2) \cup \cdots \cup P(B \cap A_n)$$

and

$$P(A_i \backslash \mathrm{B}) = \frac{P(B \backslash A_i)P(A_i)}{P(B \backslash A_1)P(A_1) + \cdots + P(B \backslash A_n)P(A_n)}$$

Developed Bayes' rule

Conditional Probability: Example

- The estimated prevalence of AIDS in a population is 10%
- We know that a diagnostic test is positive in 95% of HIV+ people, and negative un 98% of HIV- people
- What is the probability of being HIV+ if the result of the test is positive?

$$P(HIV+) = 0.1$$

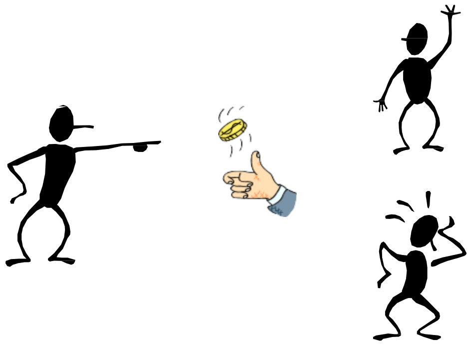
$$P(T+ \backslash HIV+) = 0.95$$

$$P(T+ \backslash HIV-) = 0.02$$

$$P(HIV+ \backslash \mathrm{T}+) = \frac{P(T+ \backslash \mathrm{HIV}+)P(HIV+)}{P(T+ \backslash \mathrm{HIV}+)P(HIV+) + P(T+ \backslash \mathrm{HIV}-)P(HIV-)}$$

$$P(HIV+ \backslash \mathrm{T}+) = \frac{0.95 \times 0.1}{0.95 \times 0.1 + 0.02 \times 0.9} = 0.84$$

Random Variable



If Tail, A wins 1 €

If Head, A loses 1 €

- Ω: {Tail, Head}
- P(Tail)=P(Head)=0.5
- G: A's gain; G=+1 if Tail; G=-1, if Head
- P(G=+1)=P(G=-1)=0.5
- G's distribution: {(+1; 0.5), (-1; 0.5)}

G is a random variable that follows a certain probability distribution

Random Variable: Definition

- Let E a set of events
- with universal finite set Ω,
- and *a* an elementary event of E

    - For any event *a* belonging to E, a number x (random variable) corresponds according to a well-defined distribution

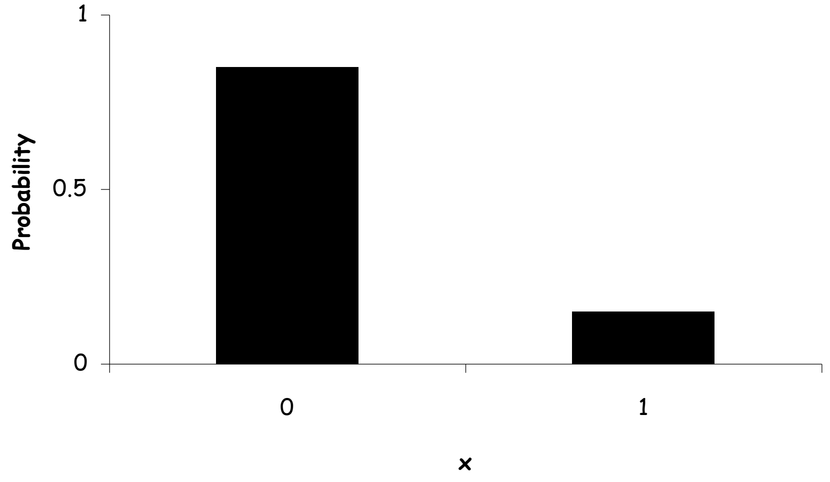
Random Variable: Example

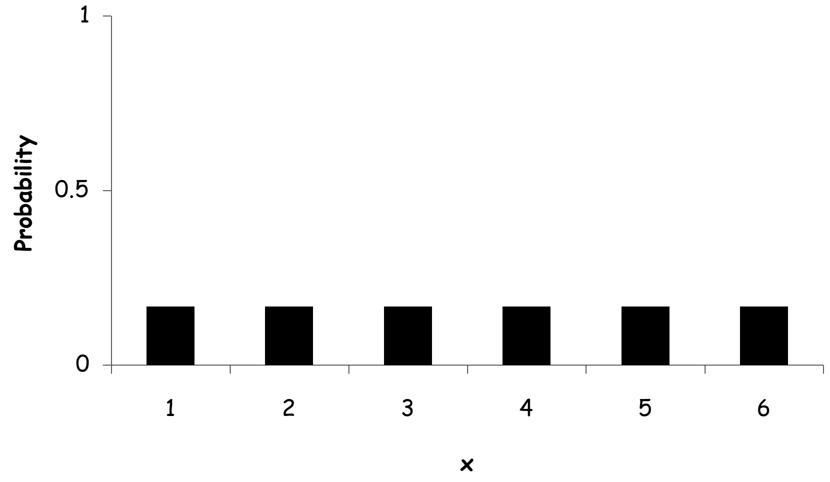
- Let disease M for which it is necessary to start treatment before the diagnosis is confirmed. However, the drug used is known to cause adverse events (AE)
- We know that: $P(M^+) = 0.05$; $P(AE^+ \backslash M^+) = 0.30$; $P(AE^- \backslash M^-) = 0.85$

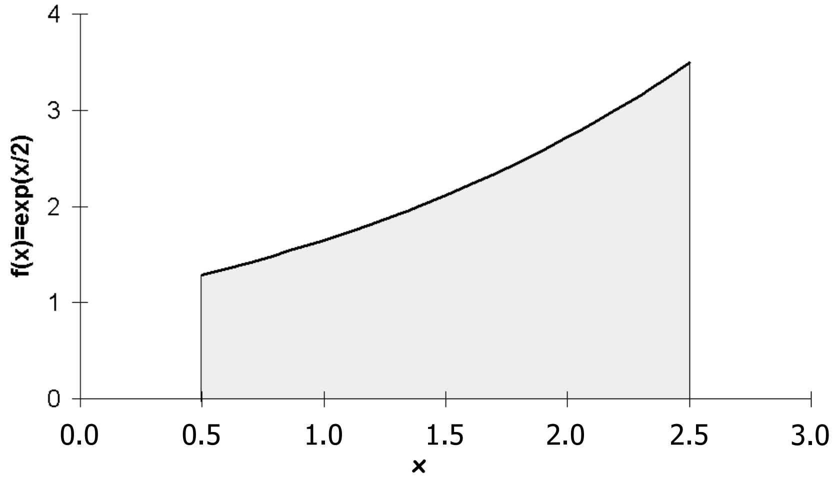
	$\boldsymbol{M^+}$	$\boldsymbol{M^-}$
$\boldsymbol{AE^+}$	$P(AE^+ \cap M^+) = 0.3 \times 0.05 = \mathbf{0.015}$ $\boldsymbol{X = 1}$	$P(AE^+ \cap M^-) = (1 - 0.85) \times (1 - 0.05) = \mathbf{0.143}$ $\boldsymbol{X = 1}$
$\boldsymbol{AE^-}$	$P(AE^- \cap M^+) = (1 - 0.3) \times 0.05 = \mathbf{0.035}$ $\boldsymbol{X = 0}$	$P(AE^- \cap M^-) = 0.85 \times (1 - 0.05) = \mathbf{0.808}$ $\boldsymbol{X = 0}$

where, X is a random variable indicator of AE.
The distribution of X is: {(0; 0.84), (1; 0.16)}

Characteristic of a Random Variable







Characteristic of Central Tendency
Mean, Mathematical Expectation

- Discrete variable X
  - Let X be a random variable taking values $x_1, x_2, \dots, x_n$ with the probabilities $p_1, p_2, \dots, p_n$ and $\sum_i p_i = 1,\ i = 1, \dots, n$

$$\mu = E(X) = \sum_i p_i x_i$$

- Continuous variable X
  - Defined by a density function $f(x)$

$$\mu = E(X) = \int_a^b x f(x) dx$$

Characteristic of Central Tendency
Mean, Mathematical Expectation

- Example 1: $\mu = (p \times 1) + \big((1-p) \times 0\big) = (p \times 1) + (q \times 0) = 0.16$

- Example 2: $\mu = 1/6 + 2/6 + 3/6 + 4/6 + 5/6 + 6/6 = 3.5$

- Example 3: $\mu = E(X) = \int_{0.5}^{2.5} f(x)dx = \int_{0.5}^{2.5} \exp(x/2)\, dx = [\exp(x/2)]_{0.5}^{2.5} = 2.21$

Characteristic of Dispersion
Variance, Standard Deviation

- Discrete variable X

$$\sigma^2 = \sum_i p_i [x_i - \mu]^2$$
$$= E\left((X-\mu)^2\right) = E(X^2) - \left(E(X)\right)^2$$

- Continuous variable X

$$\sigma^2 = \int_a^b (x-\mu)^2 f(x) dx$$

σ^2 = Variance

σ = Standard deviation

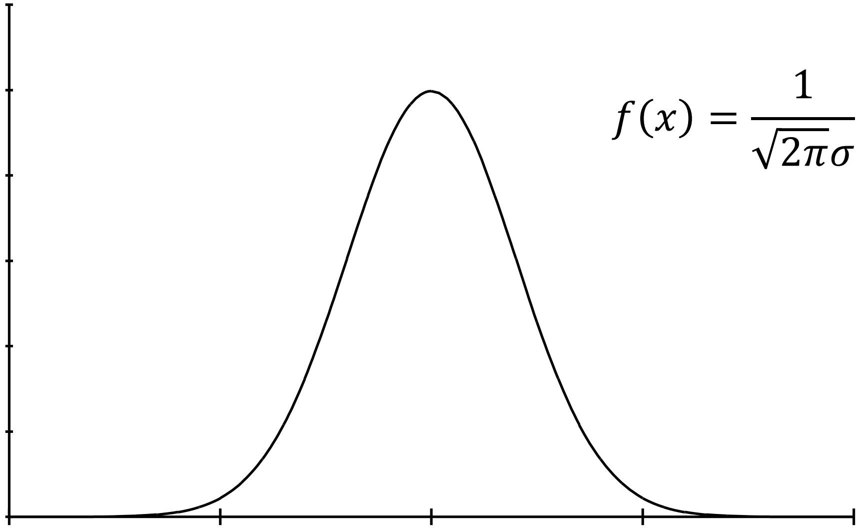
Characteristic of Dispersion
Variance, Standard Deviation

- Example 1: $\sigma^2 = p \times (1-p)^2 + q \times (0-q)^2 = pq$

- Example 2: $\sigma^2 = 1/6\,[(1-3.5)^2 + \cdots + (6-3.5)^2] = 2.9$

- Example 3: $\sigma^2 = \int_{0.5}^{2.5} (x-2.21)^2 \exp(x/2)\,dx = \int_{0.5}^{2.5} x^2 \exp(x/2)\,dx - 2.21^2 = 0.68$

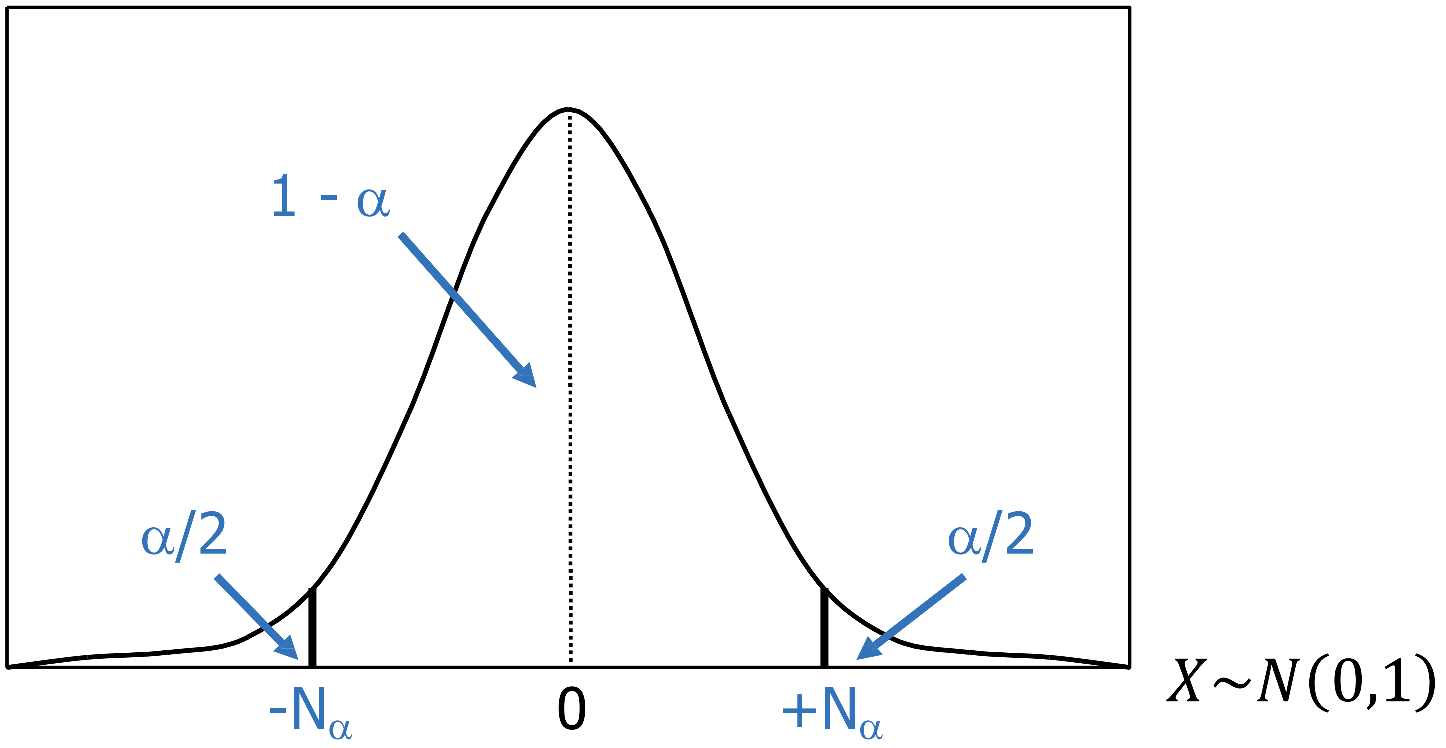
Normal Distribution (Gauss): $N(\mu, \sigma)$



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

- Properties
  - Defined by a continuous density function, determined by μ and σ
  - Density function symmetric with respect to μ
  - Density function goes to a maximum for $x = \mu$ (mode $= \mu$)
  - Median $= \mu$

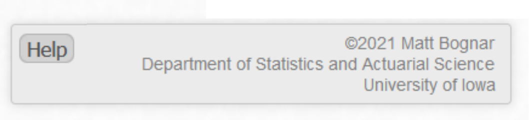
Standard Normal Distribution: $N(0,1)$

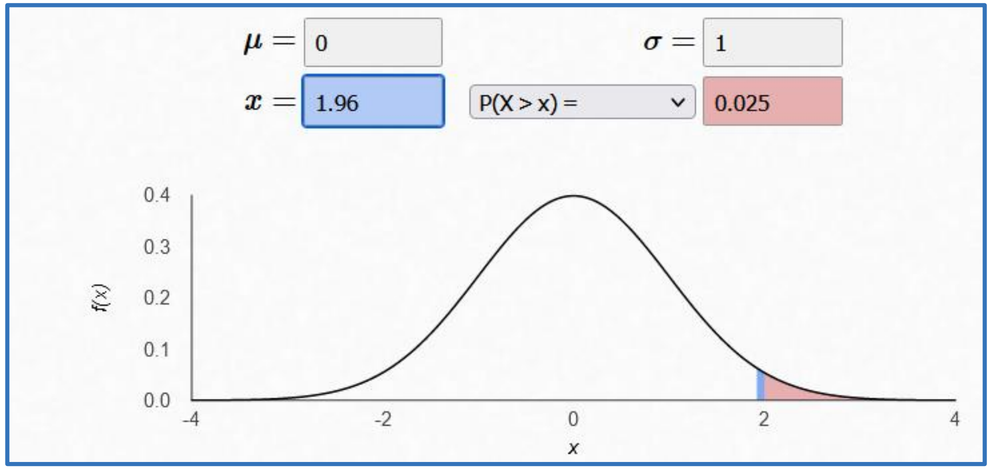


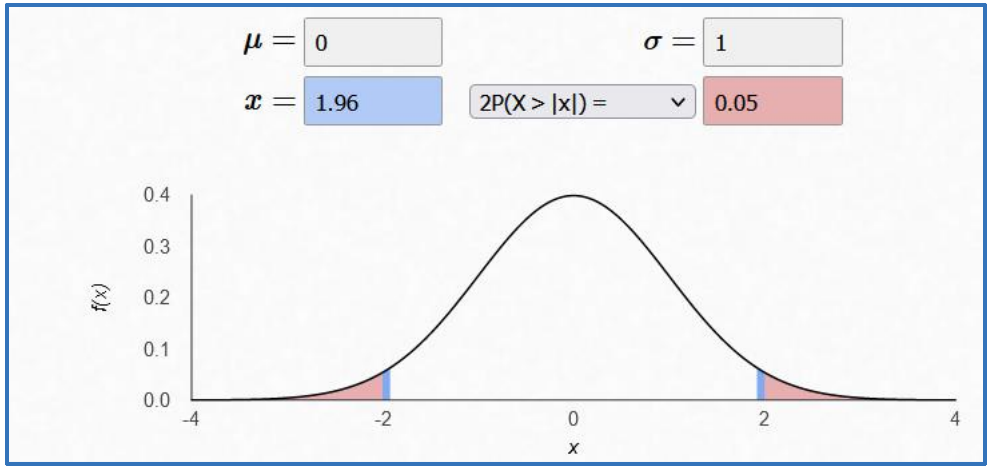
$$\alpha = P(X \leq -N_\alpha \ or \ X \geq +N_\alpha) = P(|X| \geq +N_\alpha)$$

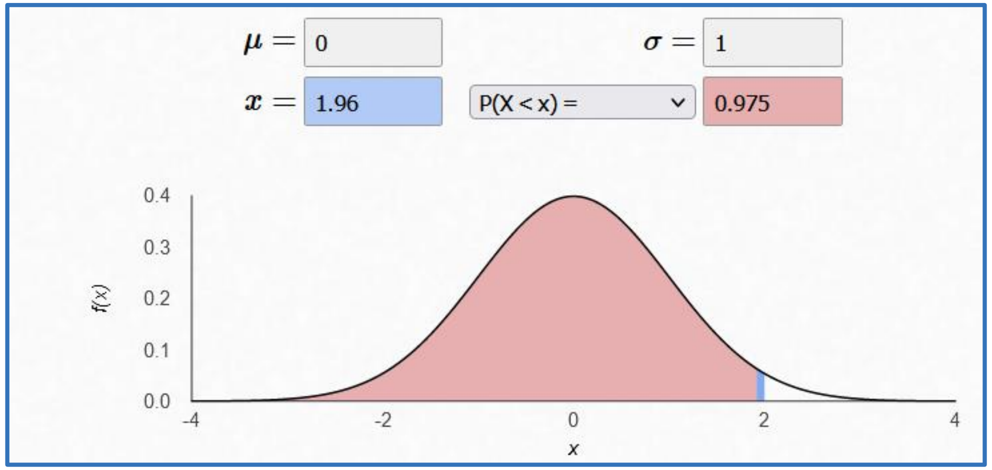
Standard Normal Distribution: $N(0,1)$

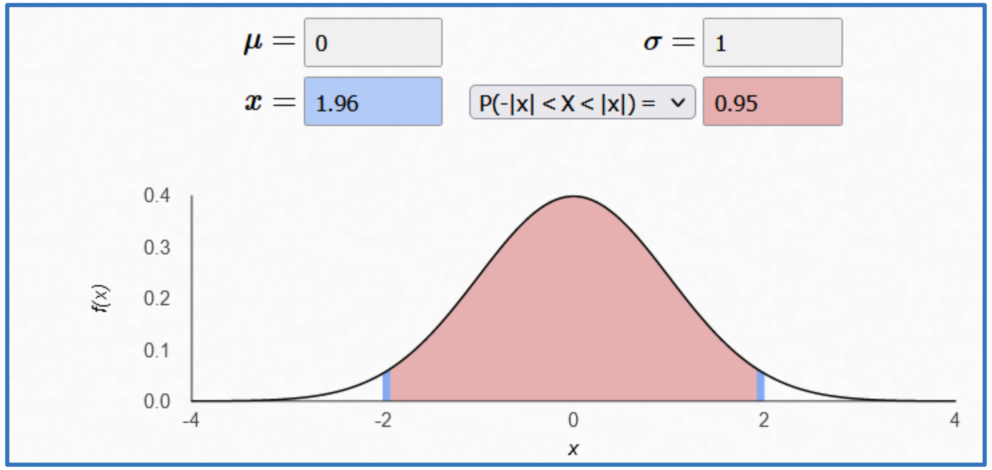
Play with: https://homepage.divms.uiowa.edu/~mbognar/applets/normal.html





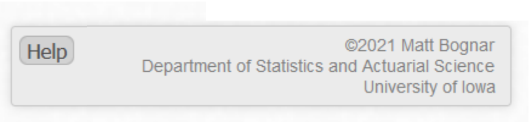


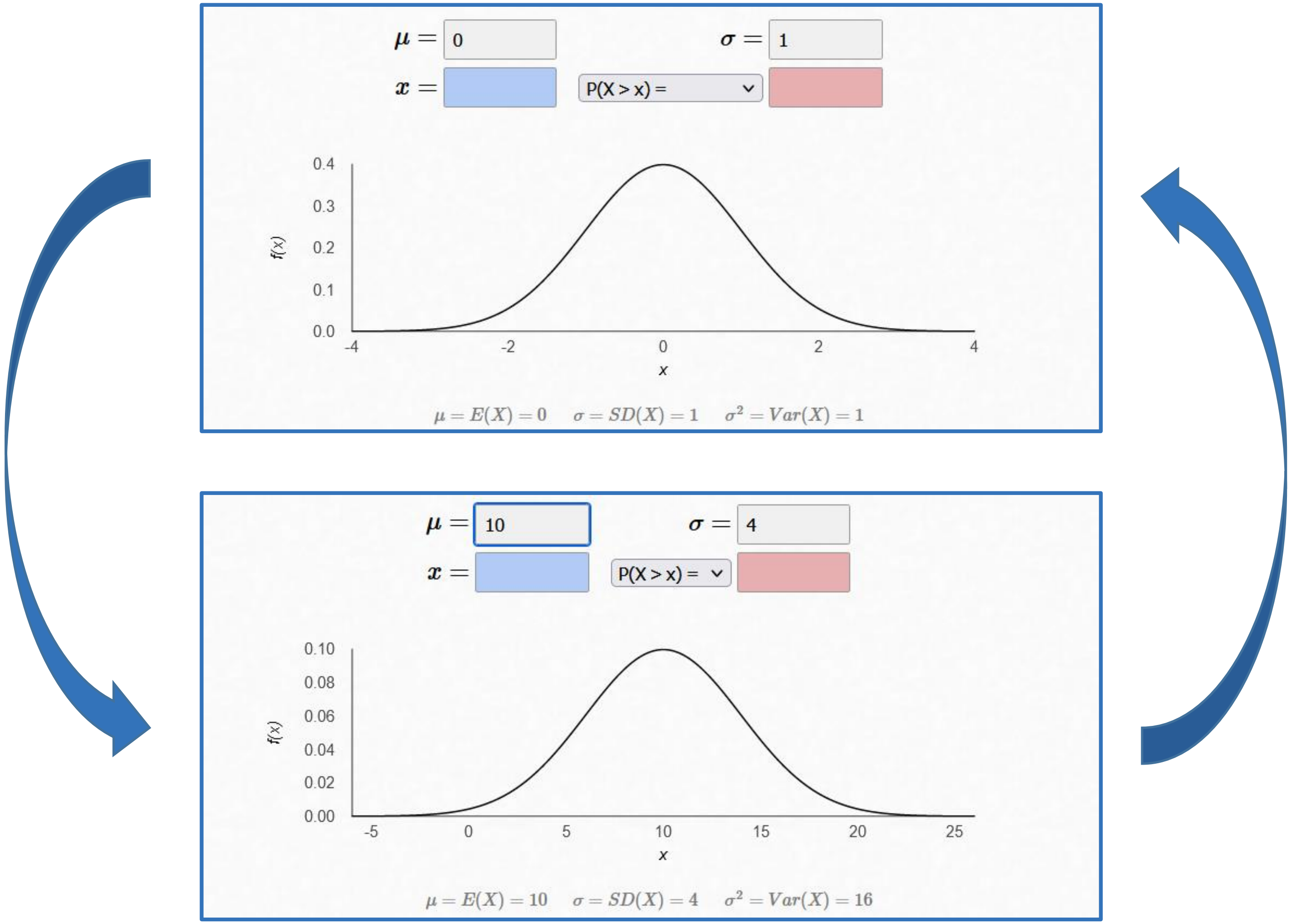




Standard Normal Distr. $N(0,1) \leftrightarrow$ Normal Distr. $N(\mu, \sigma)$

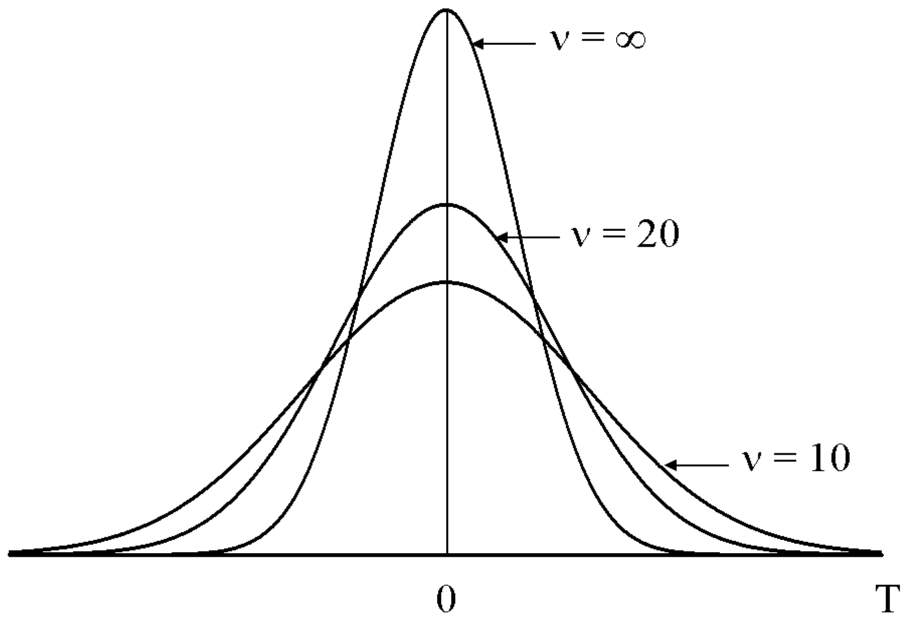
Play with: https://homepage.divms.uiowa.edu/~mbognar/applets/normal.html



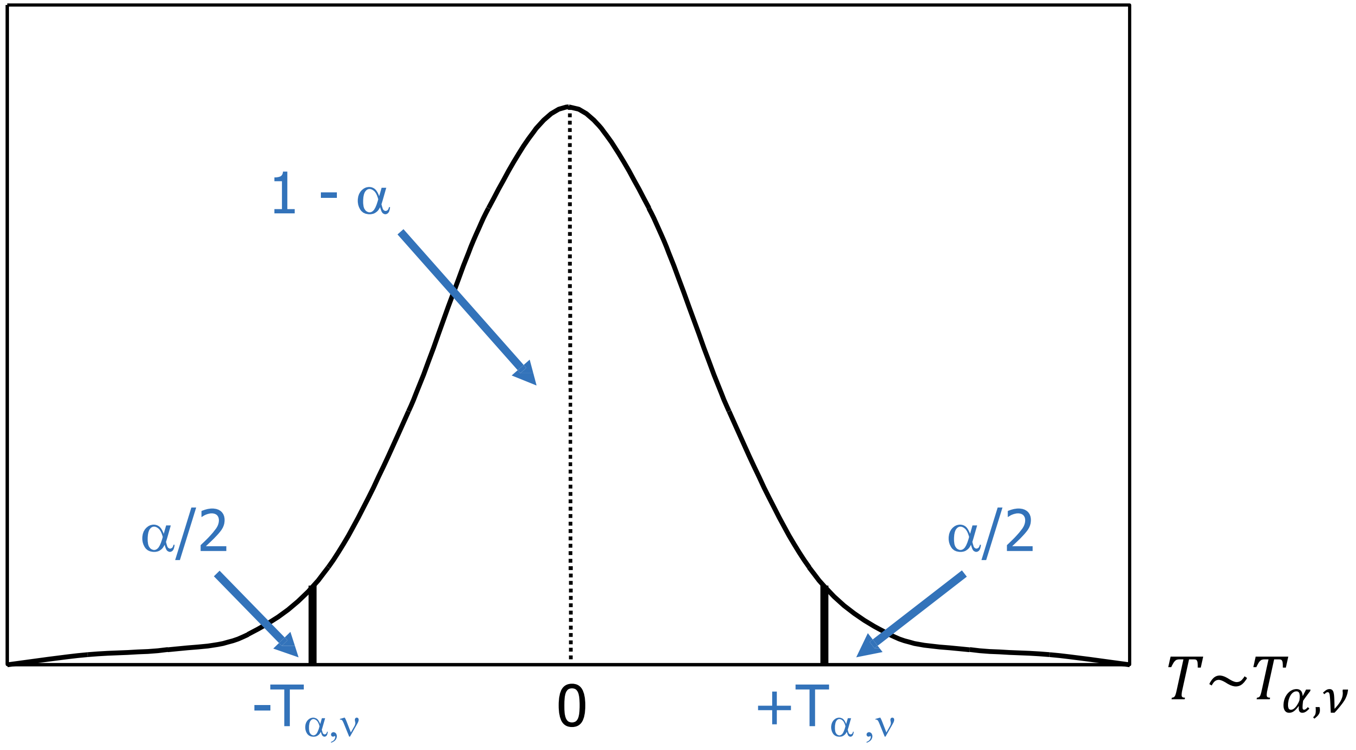


Student's t-Distribution

- ν degree of freedom (number of independents data)
- One family of Student distribution for each df
- Properties
  - Symmetric with respect to 0
  - Mode = 0
  - Flattens when ν small
  - Tends to $N(0,1)$ when $\nu \to \infty$



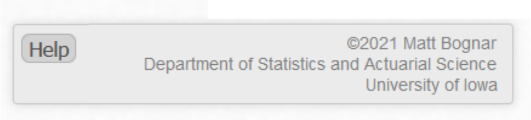
Student's t-Distribution

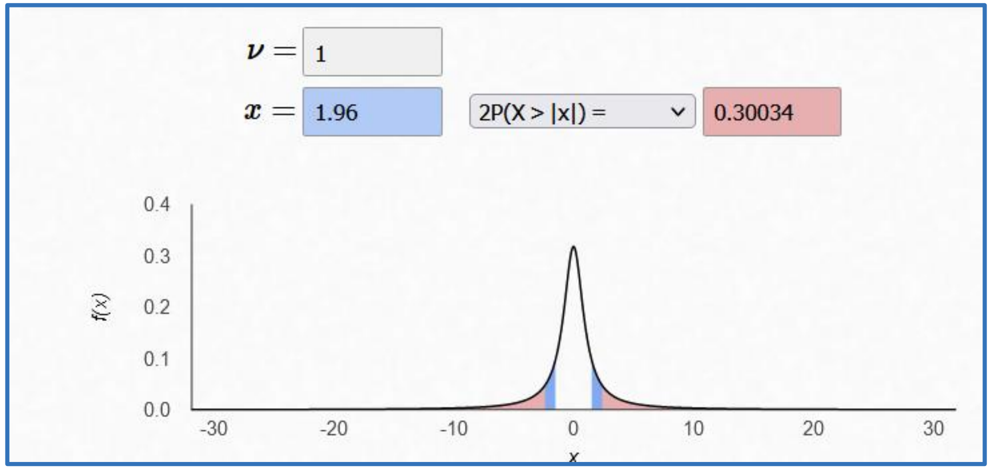


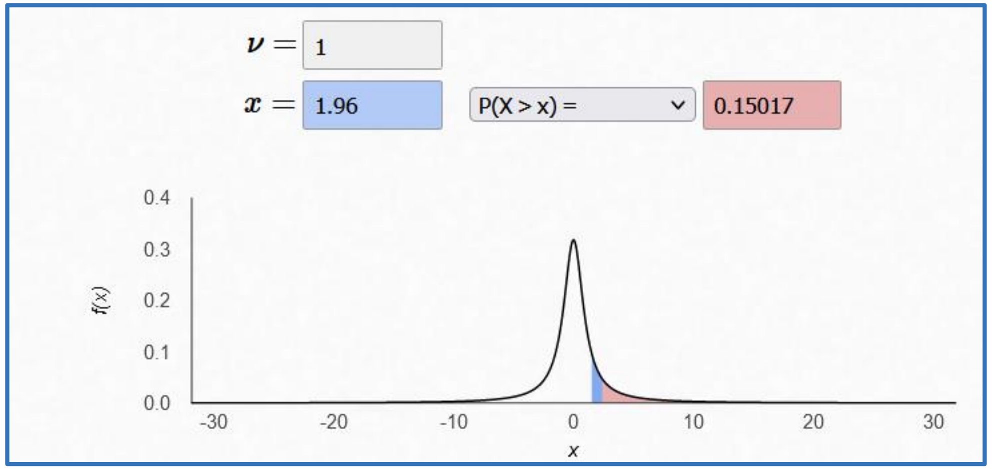
$$\alpha = P\left(T \leq -T_{\alpha,\nu} \ \ or \ \ T \geq +T_{\alpha,\nu}\right) = P\left(|T| \geq +T_{\alpha,\nu}\right)$$

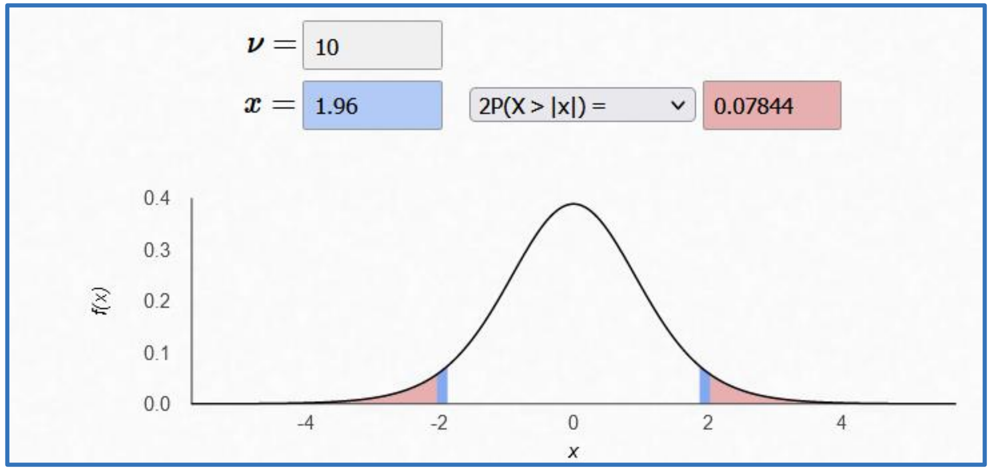
Student’s t-Distribution

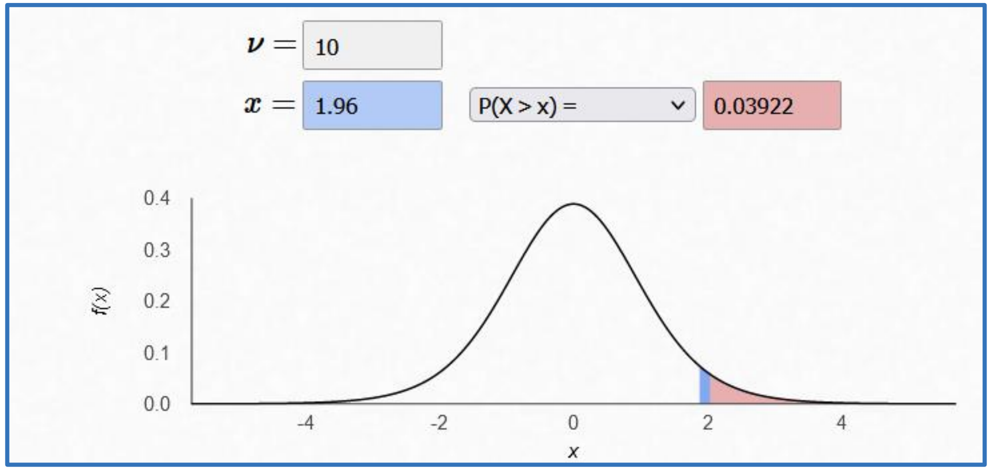
Play with: https://homepage.divms.uiowa.edu/~mbognar/applets/t.html





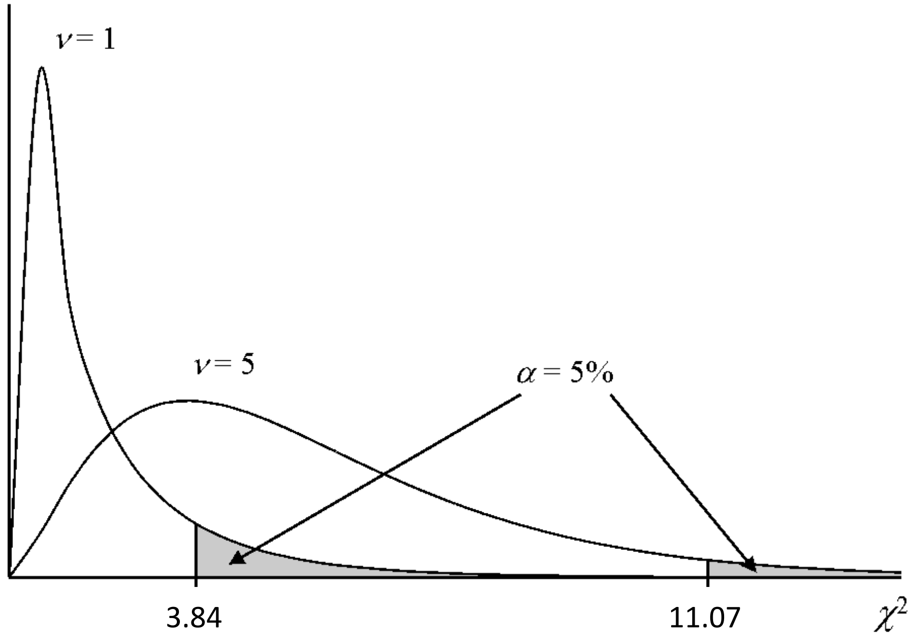






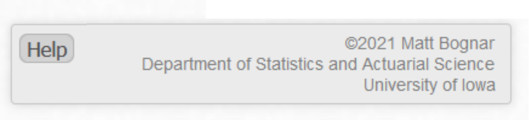
Chi-Squared Distribution

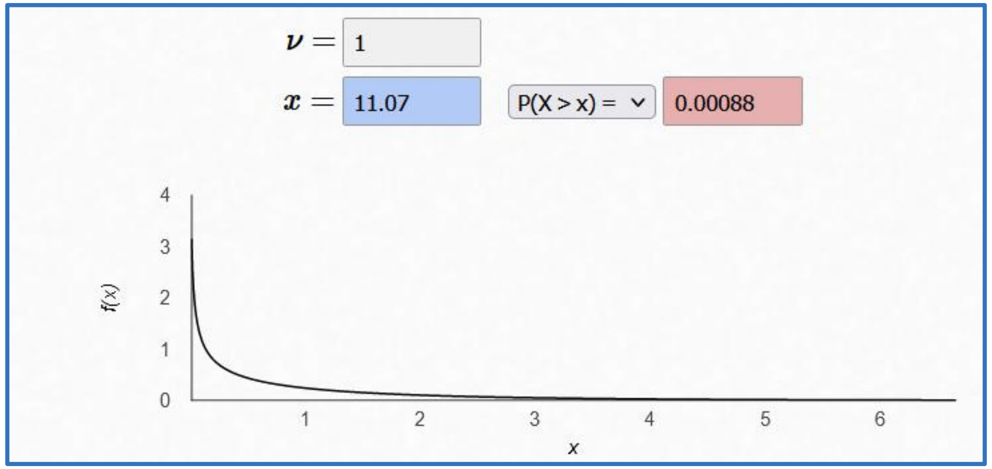
- One family of Chi-squared distribution for each df
- Properties
  - Asymmetric for ν small

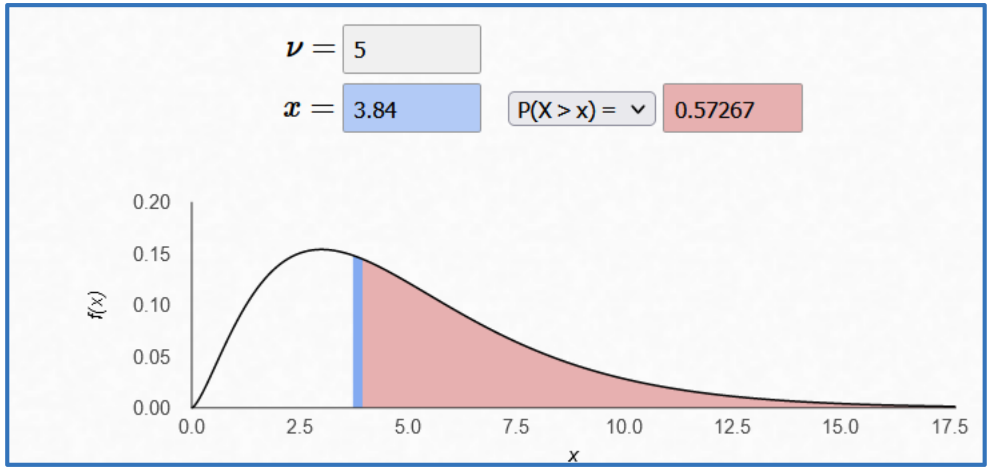


Chi-Squared Distribution

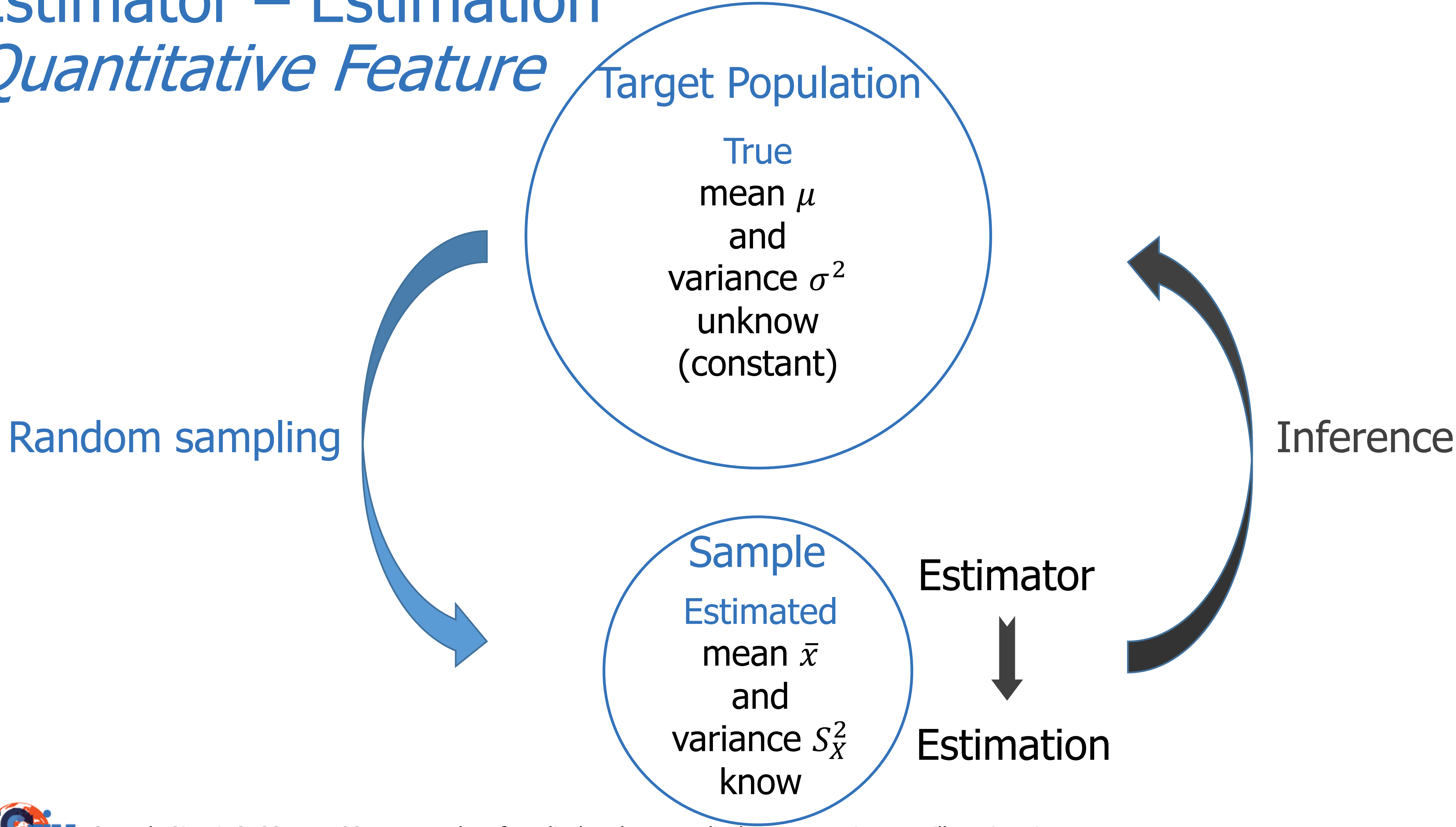
Play with: https://homepage.divms.uiowa.edu/~mbognar/applets/chisq.html





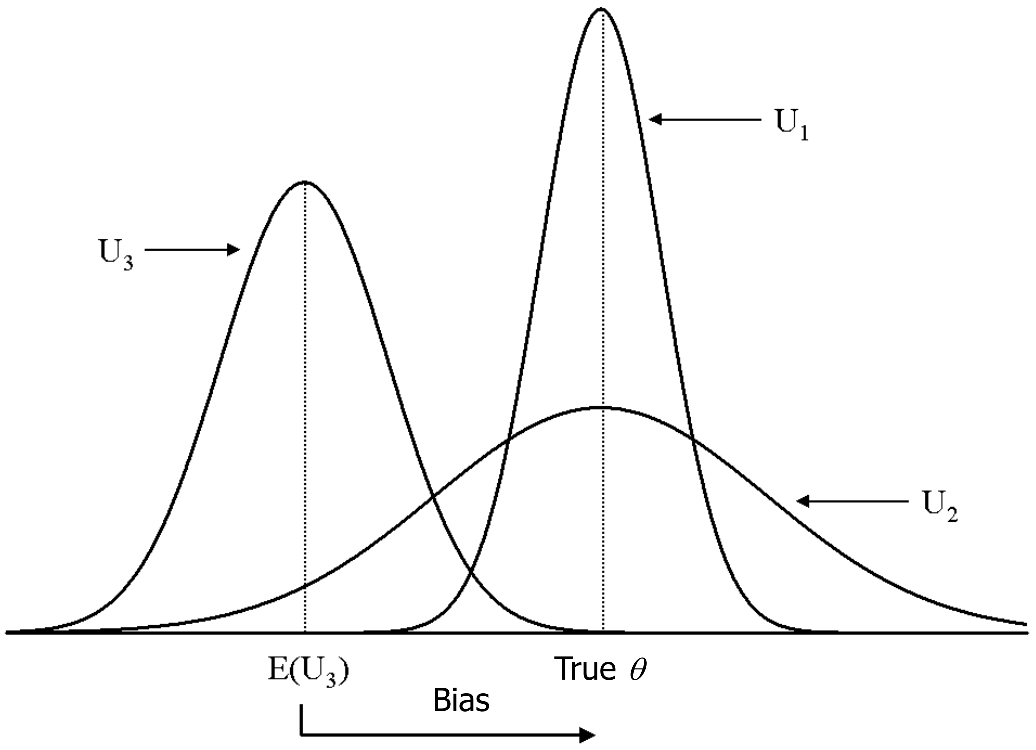


Estimator – Estimation
Quantitative Feature



Quality of an Estimator

- U: unbiased estimator of θ if $E(U) = \theta$
- U: biased estimator of θ if $E(U) \neq \theta$, and the bias = $E(U) - \theta$



Estimation of a Population Mean and Variance

Quantitative feature X, with observations $x_1, x_2, \dots, x_n$ randomly drawn from a sample of size n

- Estimator, unbiased, of μ

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

- Estimator, convergent, unbiased, of σ^2

$$S_X^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

  - $S_X = \sqrt{S_X^2}$ estimation of the standard deviation

Estimation of a Population Proportion and Variance

Qualitative feature X, with k the number of times a given characteristic is presents in a randomly sample of size n

- Estimator, unbiased, of P

$$p = \frac{k}{n}$$

- Estimator, convergent, unbiased, of $Var(P)$

$$Var(p) = \frac{p(1-p)}{n}$$

  - $\sqrt{Var(p)}$ estimation of the standard deviation

Estimation and Confidence Interval

- Estimation: point value that is desired to be close to the true (population) value of the parameter of interest (smaller or larger due to sampling fluctuations)
- Need to provide a degree (an interval) of confidence (CI) of resulting estimate
  - Determined from the data of a sample in which one can bet, with an acceptable risk of being wrong, that the true population value is

Confidence Interval

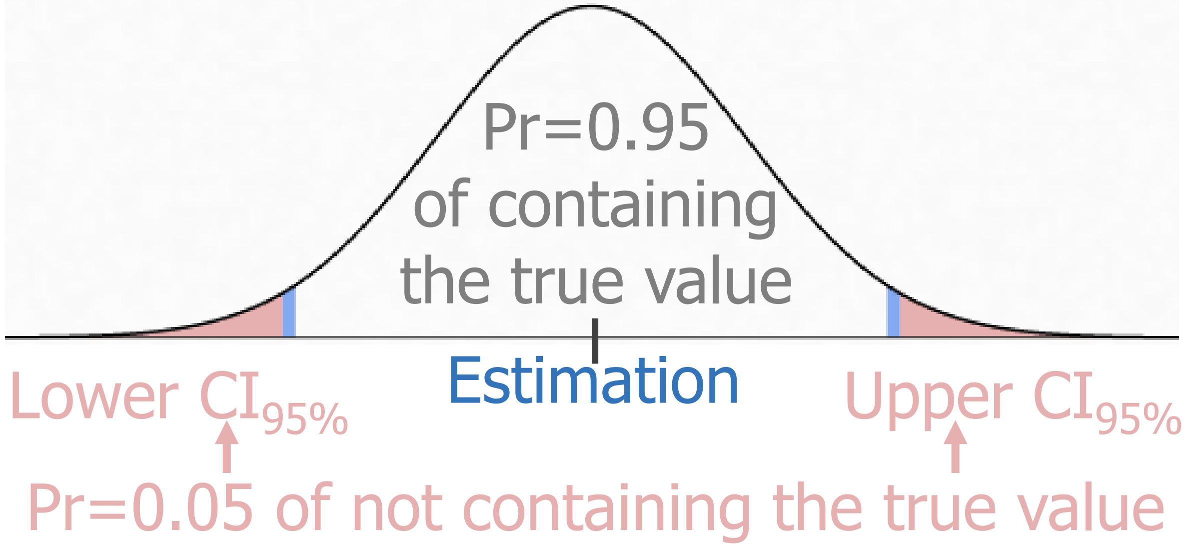
- Risk (α) corresponding to the sampling fluctuations considered as acceptable
- Confidence interval of the estimated parameter $\hat{\theta}$ has the following form

$$\hat{\theta} \text{ - sampling fluctuations ; } \hat{\theta} + \text{ sampling fluctuations}$$

- Interpretation
  - We accept that there is a α.100 chances in a hundred of being wrong by saying that the true value of the parameter of interest belongs to the interval
  - We accept that there is a $(1-\alpha)$.100 chances in a hundred of not being wrong by saying that the true value of the parameter of interest belongs to the interval

Confidence Interval

- Risk α usually=0.05
- Interpretation
  - We accept that there is a 5 chances in a hundred of being wrong by saying that the true value of the parameter of interest belongs to the interval
  - We accept that there is a 95 chances in a hundred of not being wrong by saying that the true value of the parameter of interest belongs to the interval



Confidence Interval

- Of a mean
  - If $n \geq 30$, then $L_\alpha = N_\alpha$
  - If $n < 30$, and the distribution of the feature in the population is Normal, then $L_\alpha = T_{\alpha,\nu}$

$$\bar{x} - L_\alpha \cdot \frac{S_x}{\sqrt{n}} \ ; \ \bar{x} + L_\alpha \cdot \frac{S_x}{\sqrt{n}}$$

- Of a proportion
  - If $p = k/n$ is not close to 1 or 0
  - If $p \cdot n \geq 5$ and $(1 - p) \cdot n \geq 5$

$$p - N_\alpha \cdot \sqrt{\frac{p \cdot (1-p)}{n}} \ ; \ p + N_\alpha \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$$

Statistical Tests: Example 1

- In a comprehensive cancer registry, cancer mortality in 2010 was as follows:

Lung	Colorectal	Breast	Others
24%	*14%*	*19%*	*43%*

- Has this repartition changed (effect of treatments, preventive measures,…)?
    - Unbiased survey of 1000 cancer deaths in 2020
    - Comparison of an observed distribution to a theoretical distribution

Statistical Tests: Example 1

- If the distribution has not changed at all, is there an obligation to observe?

	Lung	Colorectal	Breast	Others	Total
Reference	*24%*	*14%*	*19%*	*43%*	
Observed	240	140	190	430	1000

Statistical Tests: Example 1

- If the distribution has not changed at all, is there an obligation to observe?

	Lung	Colorectal	Breast	Others	Total
Reference	*24%*	*14%*	*19%*	*43%*	
Observed	240	140	190	430	1000

NO

▸ Sampling fluctuations

Statistical Tests: Example 1

- Do the values observed below ensure that the distribution has changed?

	Lung	Colorectal	Breast	Others	Total
Reference	*24%*	*14%*	*19%*	*43%*	
Observed	260	120	200	420	1000

Statistical Tests: Example 1

- Do the values observed below ensure that the distribution has changed?

	Lung	Colorectal	Breast	Others	Total
Reference	*24%*	*14%*	*19%*	*43%*	
Observed	260	120	200	420	1000

NO

▸ Distribution with a variability

Statistical Tests: Example 1

- Change is more likely if we have?

	Lung	Colorectal	Breast	Others	Total
Reference	*24%*	*14%*	*19%*	*43%*	
Observed	100	190	90	620	1000

Statistical Tests: Example 1

- Change is more likely if we have?

	Lung	Colorectal	Breast	Others	Total
Reference	*24%*	*14%*	*19%*	*43%*	
Observed	100	190	90	620	1000

YES

▸ No change is not impossible

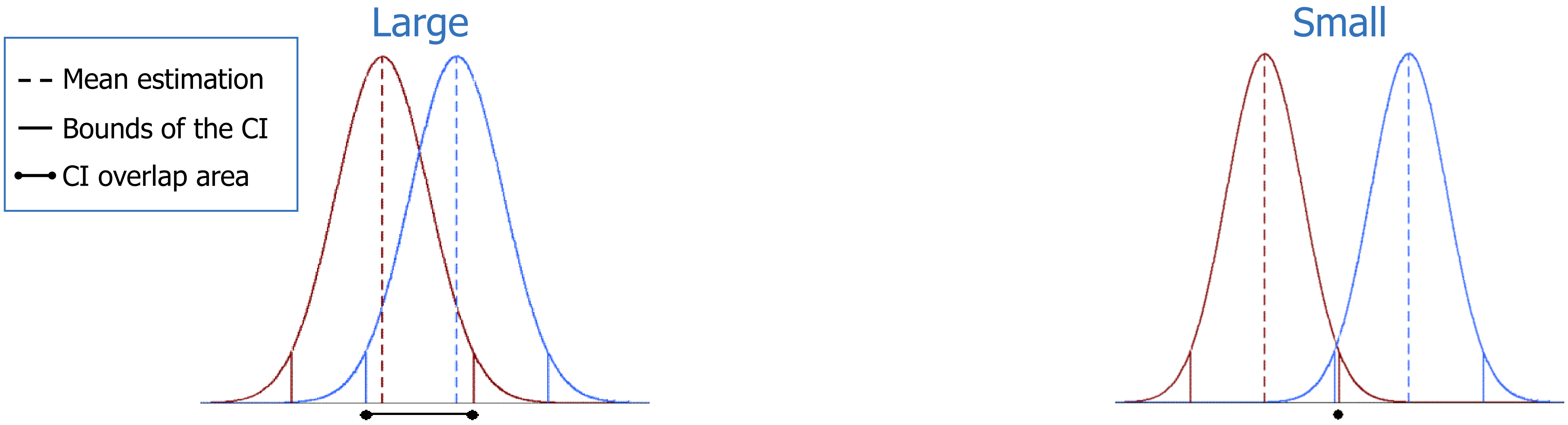
▸ More likely to change

Statistical Tests: Example 2

- Patients with atrial fibrillation (AF) treated by treatment A and patients with AF treated by treatment B
- Does the average number of AF recurrences differ according to the treatment of the first AF event?
    - Comparison of 2 independent random samples
    - Outcome: mean number of recurrences and confidence intervals from an unbiased sample of patients

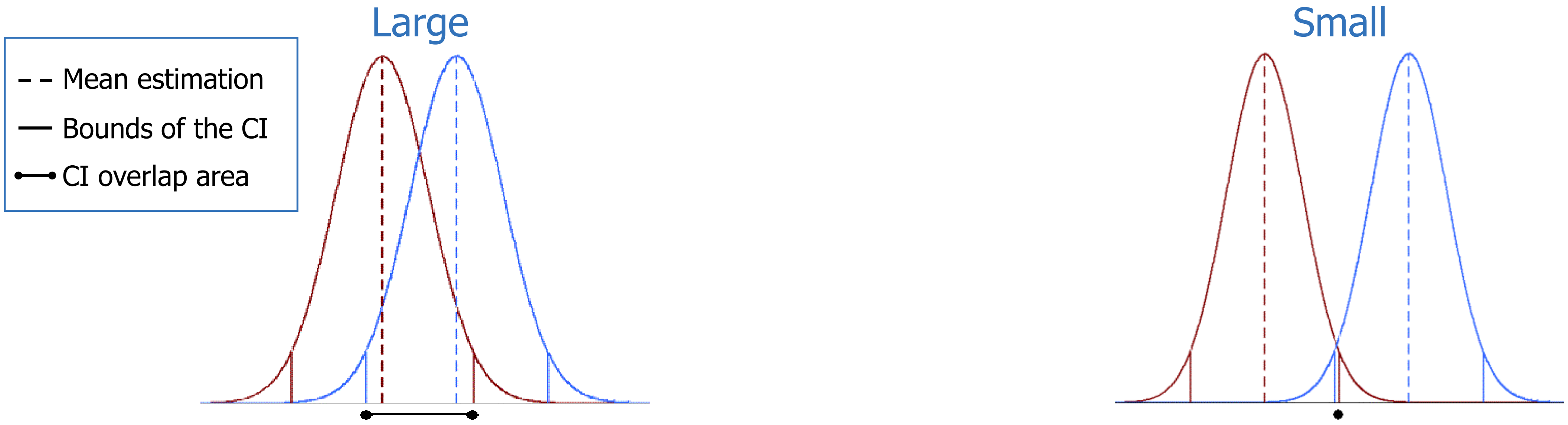
Statistical Tests: Example 2

- The overlap area of the confidence interval (CI) is



Statistical Tests: Example 2

- The overlap area of the confidence interval (CI) is



- The observed difference between the estimated means is due to sampling fluctuations
- The 2 samples come from the same population

- It is “unlikely” that the observed difference between the estimated means is due to chance
- The probability that the 2 samples come from the same population is “low”

Statistical Tests: Principles

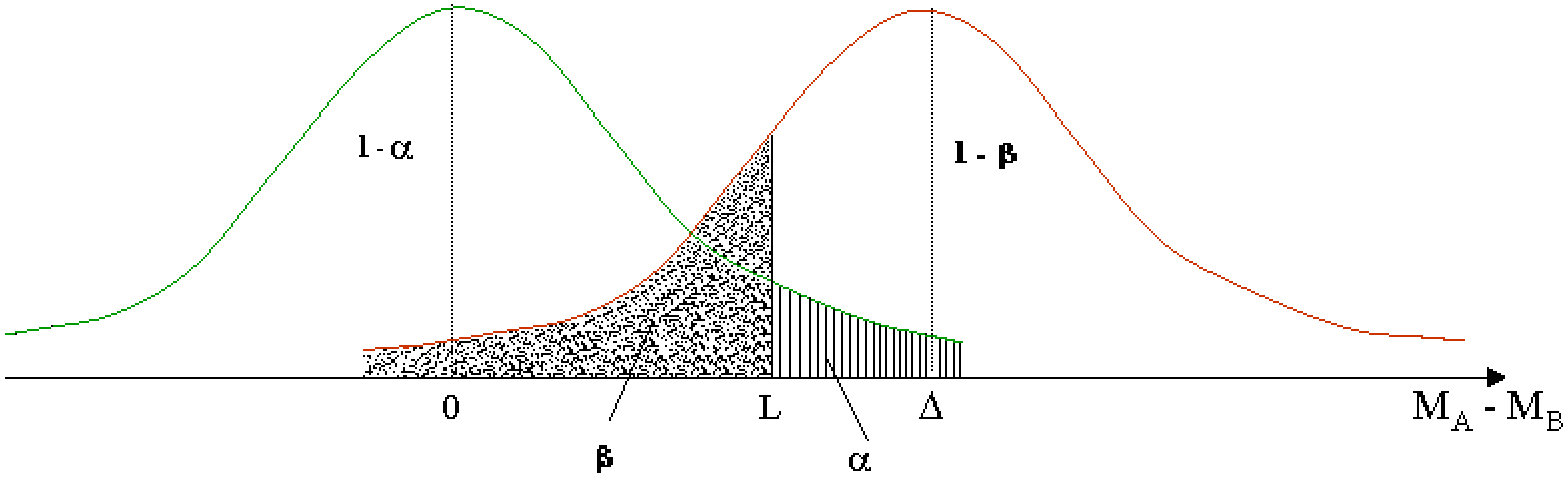
- How to define a threshold between a “large” and “small” coverage area?
  - Value that satisfied the hypothesis that the estimated difference between the means is due to sampling fluctuations: null hypothesis (H_0)
  - This hypothesis will be rejected if the difference between the estimated distribution and the theoretical distribution is too "large", i.e. H_0 is too implausible

Statistical Tests: Example 2

- Estimated mean number of recurrences
  - $\bar{x}_A$ in Group A
  - $\bar{x}_B$ in Group B
- Hypothesis
  - Null H_0: $\mu_A = \mu_B$, there is no treatment difference
  - Alternative (two-sided) H_1: $\mu_A \neq \mu_B$, there is a treatment effect
- Statistic of the appropriate test, $d = M_A - M_B$
- A significance level L is determined at which $|d|$ is considered to be "too large", i.e. such that
  - $\alpha = P[|d| > L$ under the null], usually the level of significance $\alpha = 0.05$
- Statistical test of difference between two population means
  - If the obtained p-value is less than the significant level we reject the null hypothesis, in other case we do not reject the null hypothesis

Statistical Tests: Types of Errors

Distribution of $M_A - M_B$ under H_0 Distribution of $M_A - M_B$ under $H_1: \mu_A - \mu_B = \Delta$



$\beta = P[|M_A - M_B| < L \text{ under } H_1: \mu_A - \mu_B = \Delta]$

$\alpha = P[|M_A - M_B| > L \text{ under } H_0]$

- Notations
  - $M_A - M_B$: random variables
  - $\mu_A - \mu_B$: theoretical differences

Statistical Tests: Types of Errors

- The decision to reject or not H_0 is made under certain errors, since the true state is unknown

		Reality (unknown)	
		H_0 is true	H_1 is true
Decision from the statistical test result	Don't reject H_0	Correct inference Probability = $1 - \alpha$	Type II error Probability = β
	Reject H_0	Type I error Probability = α	Correct inference Probability = $1 - \beta$ Statistical Power

Statistical Tests: Interpretation

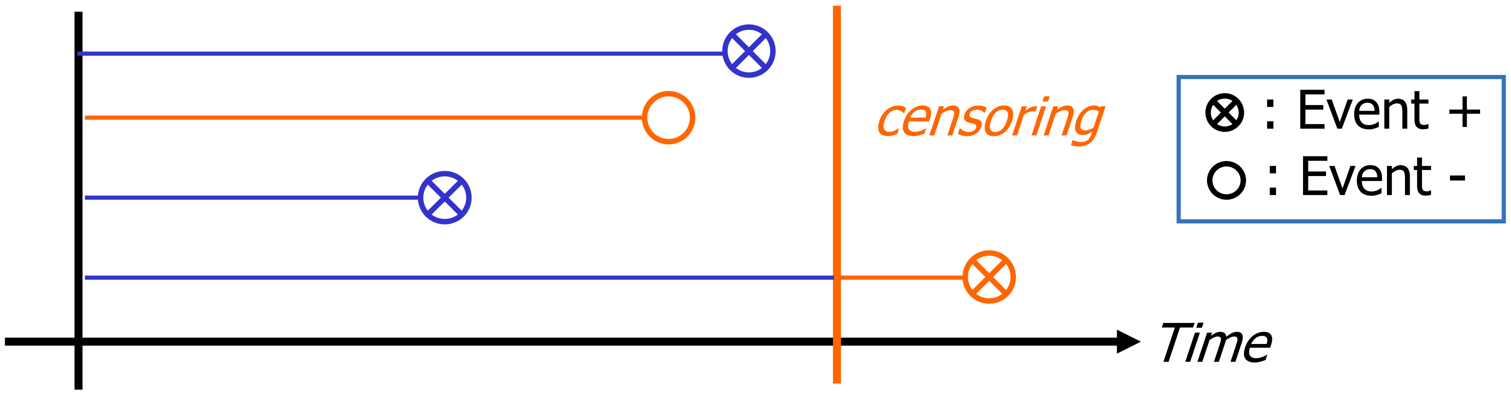
- P-value
  - Probability of obtaining test results as least as extreme as the results actually observed, under the assumption that the null hypothesis is correct
  - $p = P[|M_A - M_B| > d$ under $H_0]$
  - Information in terms of probability of the distance between the observed value of the statistic and an expected value under H_0
  - Does not measure the strength of an effect (i.e. means difference, relative risk,…)
- Interpretation of a statistical test
  - $p > \alpha$: non rejection of H_0 (non significant result –in a statistical point of view)
  - $p \leq \alpha$: rejection of H_0, with error of type I (significant result –in a statistical point of view)

Statistical Tests: Choice

- Type of the features
  - Qualitative × Qualitative
    - Sex × Tumour stage
  - Qualitative × Quantitative
    - Sex × Tumour size (mm)
  - Quantitative × Quantitative
    - Biological marker (UI/mL) × Tumour size (mm)
- Type of samples
  - Unpaired (independent): distinct "subjects"
  - Paired (dependent): same "subjects"
- Test hypotheses and methodological hypotheses of the methods

Statistical Tests: Choice

- Censored data or not-censored data
  - Censored: patient follow-up ended before the event of interest occurred
    - Event: death, recurrence of a disease,…



Statistical Tests: Choice

- Univariate or multivariate statistical analysis
  - Univariate: the event of interest is explained only by a single feature
    - Pharyngeal cancer = f(age)
    - Pharyngeal cancer = f(smoking)
    - Pharyngeal cancer = f(alcohol)
  - Multivariate: the event of interest is explained by several features taken together
    - Pharyngeal cancer = f(age, smocking, alcohol)

Univariate Analysis – Not-Censored Data

	Qualitative	Quantitative
Qualitative	Chi-squared	Means comparison * Analysis of variance
Quantitative		Correlation coefficient + Simple linear regression

* • "Large" samples: Student's t-test (paired or unpaired)
- "Small" samples: nonparametric tests
  - Unpaired: Mann-Whitney U-test, Kruskal-Wallis test
  - Paired: Wilcoxon test, Friedman test

\+ • "Large" samples: Pearson's correlation
- "Small" samples: Spearman's correlation

Univariate Analysis – Censored Data

- Survival analysis
  - Kaplan-Meier estimator
- Comparison of survival distributions
  - Log-Rank test

Multivariate Analysis – Not-Censored Data

Dependent Variable	Predictor features	Method
Qualitative	Qualitative or Quantitative	Logistic regression, multinomial regression
Quantitative	Qualitative or Quantitative	Multiple linear regression

Multivariate Analysis – Censored Data

- Survival analysis
  - Cox proportional hazards model (qualitative of quantitative features)