Bioestatística

M.I. Eng. Biomédica

2015-2016

Aula Teórica 5

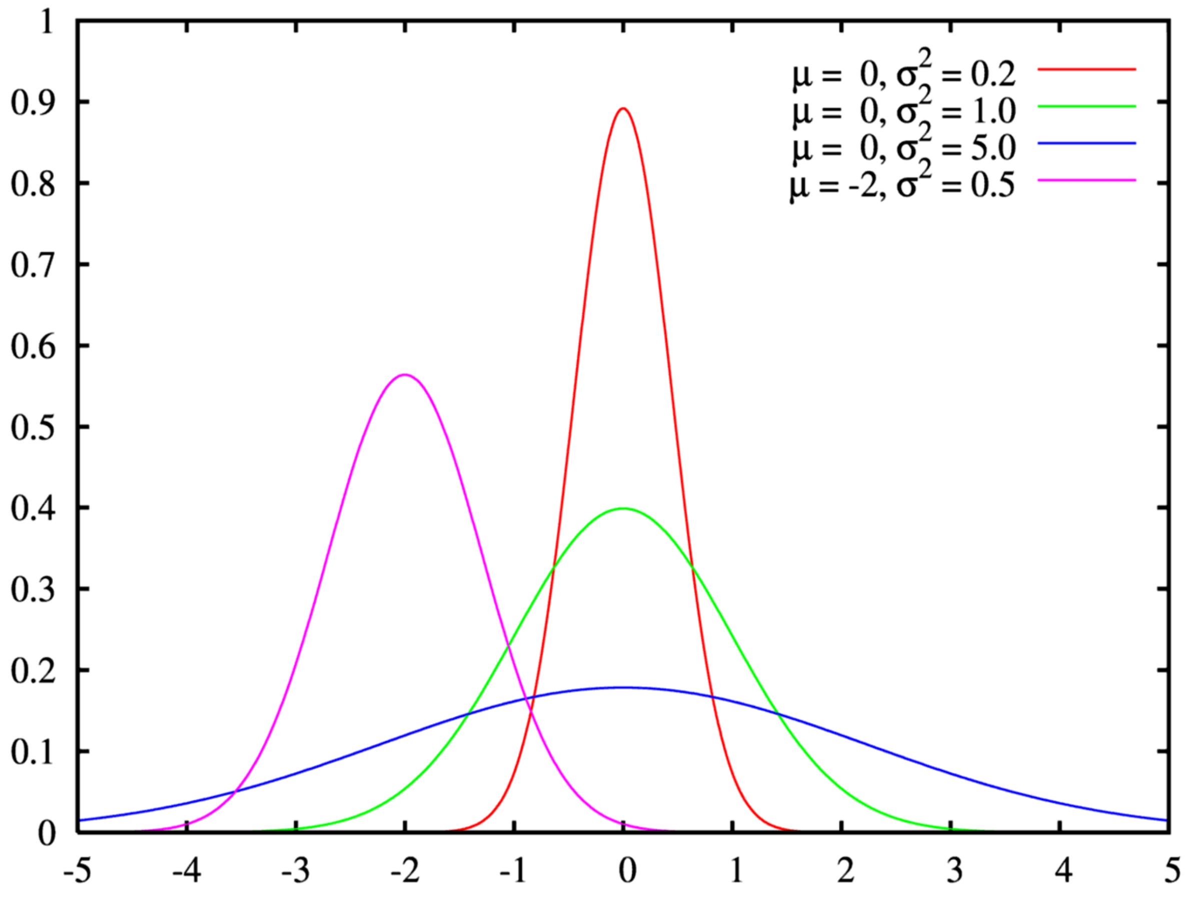
Normal

- Definição
  - Diz-se que uma variável aleatória X tem distribuição normal com parâmetros μ e σ^2 quando a função densidade é da forma

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

$$X \sim N(\mu,\sigma^2)$$

Normal



http://en.wikipedia.org/wiki/File:Normal_distribution_pdf.png

Normal

- Propriedades
  - $E(X) = \mu$;
  - $Var(X) = \sigma^2$

Normal

- Exemplo
  - A precipitação anual (em mm) no distrito de Beja é bem modelada por uma distribuição normal com $\mu = 572\, mm$ e $\sigma = 138.6\, mm$. Qual é a probabilidade da precipitação anual se situar entre 700 e 800 mm?

$$P(700 < X < 800) = \int_{700}^{800} f(x)\,dx = F(800) - F(700)$$

$$P(700 < X < 800) = \Phi\left(\frac{800-572}{138.6}\right) - F\left(\frac{700-572}{138.6}\right) = 0.1279$$

Normal

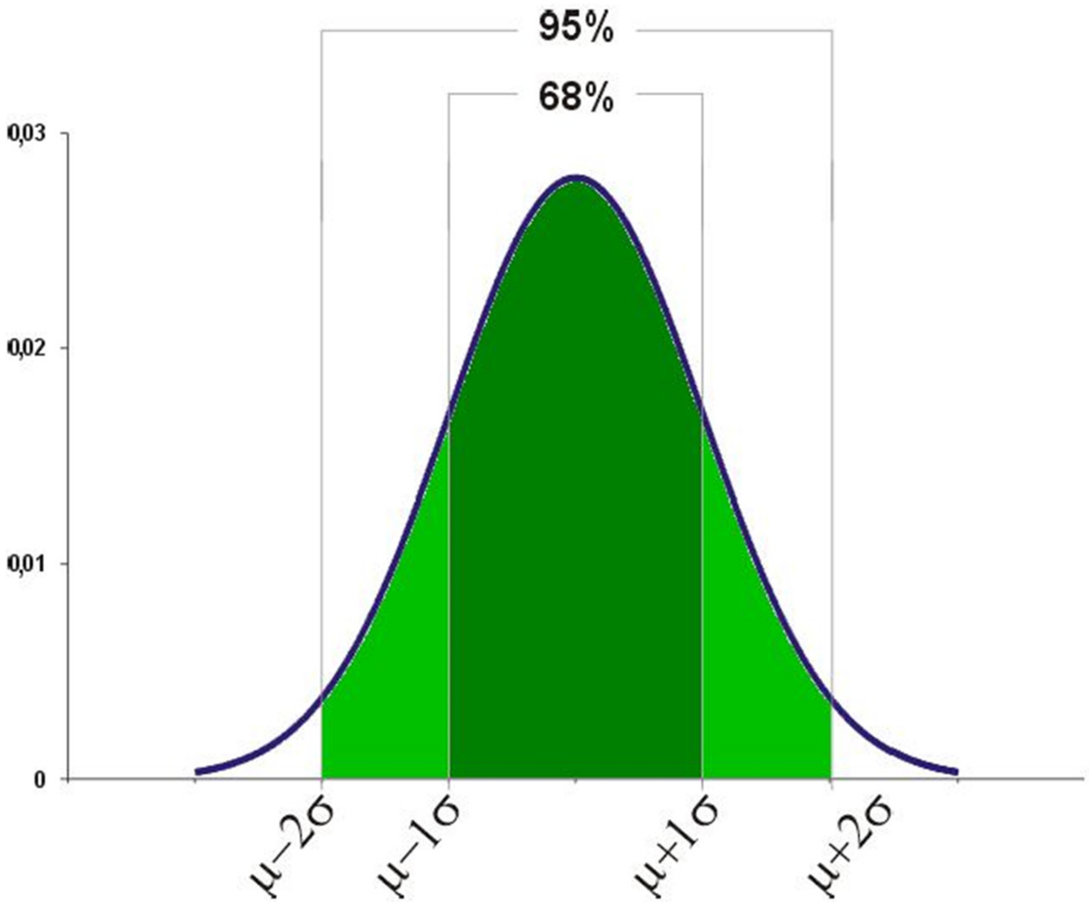
- Tabela – N(0,1)

Normal Deviate z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-4.0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.9	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.8	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.7	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

Normal

Intervalos	Probabilidades
$\mu \pm \sigma$	0.6826
$\mu \pm 2\sigma$	0.9544
$\mu \pm 3\sigma$	0.9973
$\mu \pm 0.6745\sigma$	0.5000
$\mu \pm 1.6450\sigma$	0.9000
$\mu \pm 1.9600\sigma$	0.9500
$\mu \pm 2.5758\sigma$	0.9900

Normal



Inferência

- Na teoria da probabilidade
  - Modelo probabilístico → probabilidade de resultados

> A probabilidade de obter um falso negativo com um teste de gravidez é de 0.01. Qual a probabilidade de em 100 mulheres grávidas testadas haver no máximo 2 testes falsos negativos?

- Na inferência estatística
  - Dados/observações → modelo

> A empresa SAIBAJÁ lançou um novo teste de gravidez. Para afeirir da percentagem de falsos negativos decidiu testar o produto em 100 mulheres grávidas.

Inferência

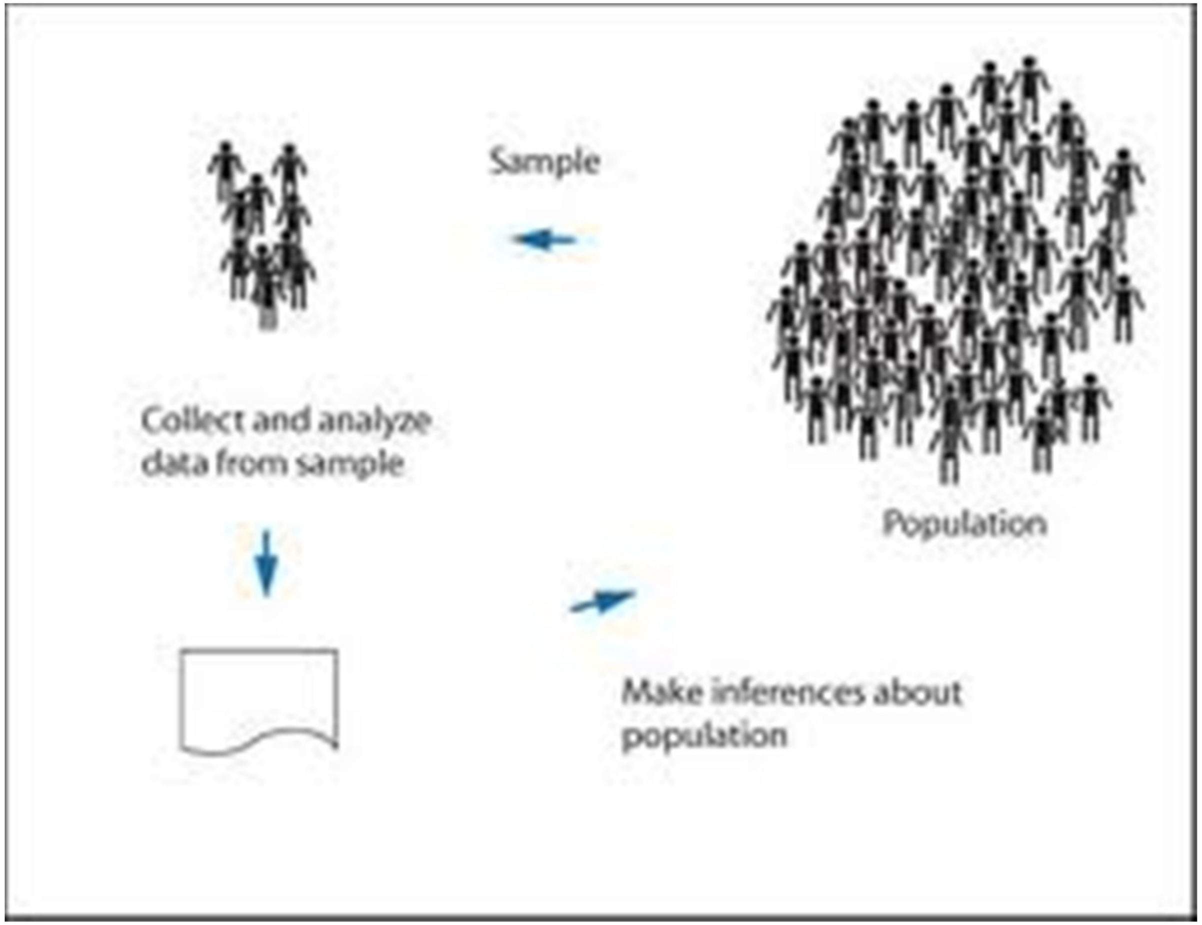
- O observador pode ter interesse em:
  - estimar a partir da observação da amostra a proporção de testes falsos negativos;
  - construir, com base na amostra, um intervalo que, com razoável confiança, contenha o valor desconhecido dessa proporção;
  - proposta a hipótese de que a proporção de testes falsos negativos é inferior a 1%, averiguar em que termos os dados suportam essa proposição;

Inferência

- A inferência é um problema central na estatística:
  - partindo de um conjunto de dados pretende-se 'determinar' as propriedades da distribuição que está na sua base.

- A inferência estatística é usualmente dividida em
  - Estimação
    - parâmetros específicos da população
  - Decisão (testes de hipóteses)
    - decisão sobre o valor de um determinado parâmetro da população;

Inferência


Sample
Population
Collect and analyze data from sample
Make inferences about population

Amostragem

- Censo
  - informação relativa a todos os elementos da população;

- Amostragem:
  - analisa-se um subconjunto da população

Amostragem

- Vantagens da amostragem
  - impossível a recolha de todos os elementos da população em:
    - populações infinitas ou com elevado nº de elementos;
    - quando o estudo das características de cada elemento conduz à sua destruição;
  - O estudo cuidadoso de uma amostra conduz a resultados mais fidedignos do que o estudo sumário de toda a população;
  - Menor custo e obtenção de resultados em tempo oportuno;
  - Problemas de ordem ética devem ser tidos em consideração:
    - estudo de novos medicamentos ou de novas técnicas cirúrgicas;
    - técnicas invasivas

Amostragem

- Amostras de conveniência
  - são, muitas vezes, as únicas possíveis de obter, principalmente quando se trata de populações raras, mal conhecidas, geograficamente mal determinadas;
  - perigo de tendenciosidade, logo inadequadas para produzir inferência;
- Amostragem aleatória, casual ou probabilística
  - é a que garante melhor representatividade;
  - é necessário possuir uma listagem de todos os elementos da população de modo a que a probabilidade de qualquer elemento da população ser seleccionado seja conhecida à priori ($\neq 0$.)
  - extremamente difícil obter tal amostragem, mas possível obter uma aproximação

Amostragem

- Amostragem aleatória
  - Simples:
    - todos os elementos têm igual probabilidade de serem seleccionados (1/N) por sorteio;
    - este método não é muito usado dado que é difícil obter populações réplica;

Amostragem

- Amostragem aleatória
  - Estratificada
    - quando se conhece a estrutura da população,
    - conduz a amostras representativas de menor dimensão;
    - a população é dividida em estratos, grupos homogéneos relativamente a uma característica (ex: sexo), e dentro de cada estrato seleccionam-se os elementos duma forma aleatória simples, de acordo com a proporção de cada grupo na população;

Amostragem

- Amostragem aleatória
  - Sistemática ou quase aleatória
    - apenas o 1º elemento da amostra é escolhido aleatoriamente, e os restantes são determinados de modo sistemático pela razão N/n (N – dimensão da população; n – dimensão da amostra);
    - o 1º elemento pode ser obtido por uma tabela de nºs aleatórios no intervalo [1, N/n], e os restantes por adição de N/n (valores arredondados ao menor inteiro);

- Amostragem
  - Conveniência
    - ✓ Por vezes, única forma
    - ✓ Enviesamentos
  - Aleatória
    - Simples
    - Estratificada
    - Sistemática
    - ✓ Melhor representatividade
    - ✓ Difícil de obter
  - ✓ Tamanho da população
  - ✓ Destruição da população
  - ✓ Menor custo
  - ✓ Menos tempo
  - ✓ Problemas éticos

Estimação

- O objectivo da estimação é estimar parâmetros de uma população teórica a partir de estatísticas obtidas numa amostra representativa dessa população.
- Se se extraírem n amostras de uma população cuja função de probabilidade (densidade) depende de um parâmetro (e.g. a média: μ) do qual se desconhece o verdadeiro valor, é necessário estimá-lo, com um determinado grau de
  - precisão (estimação por pontos)
  - confiança (estimação por intervalos)

Estimação

- Um estimador natural para estimar a média , μ, de uma população é a média amostral:

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

- O melhor estimador do desvio padrão:

$$s^* = \sqrt{\frac{n-1}{n}}\, s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Estimação

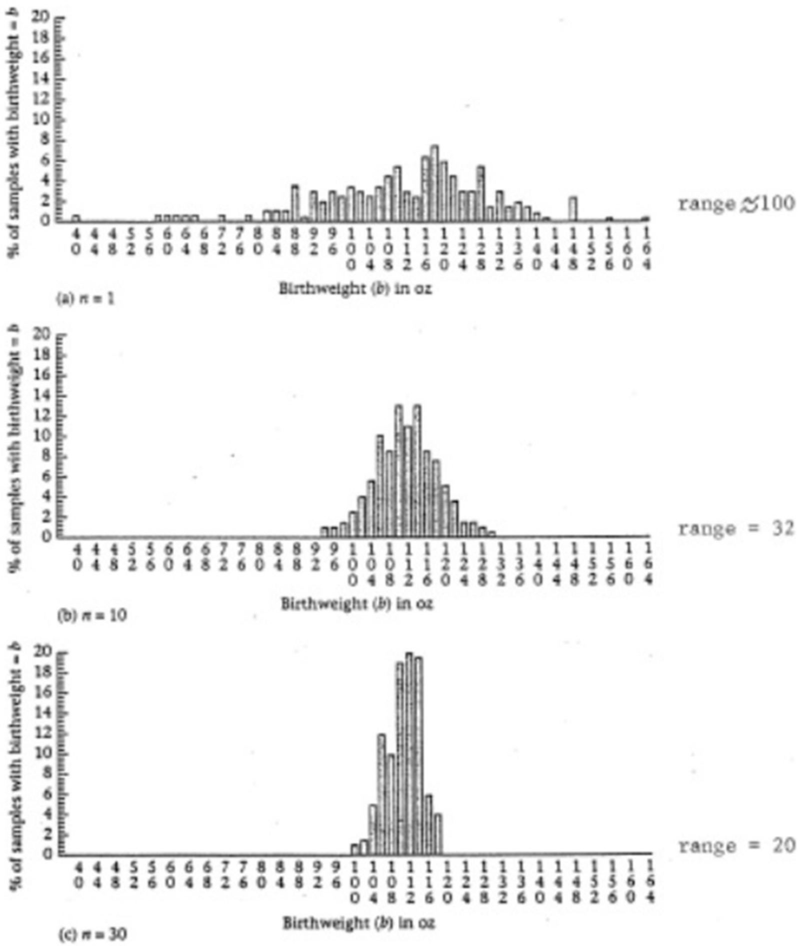
- Considere-se todas as possíveis amostras com tamanho *n* que se poderiam retirar da população;
  - As médias obtidas para cada uma dessa amostras seriam, previsivelmente, diferentes;
  - Assim, a amostra 'colhida' deve ser tomada como representativa de todas as amostras (de tamanho *n*) possíveis;

- Seja $X_1, X_2, \cdots, X_n$ uma amostra casual de população para a qual existe média μ. Então para a média amostral $\bar{X}$, $E(\bar{X}) = \mu$.

Estimação

- A média amostral é um estimador da média da população qualquer que seja o tamanho da amostra.
- Quanto maior o tamanho da amostra melhor a estimativa?

Estimação



N = 1

N = 10

N = 30

Estimação

$$Var(\bar{X}) = Var\left\{\sum_{i=1}^{n} \frac{X_i}{n}\right\} = \frac{1}{n^2}\, Var \sum_{i=1}^{n} X_i = \frac{1}{n^2}\,(n\,\sigma^2)$$

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

Estimação

- O erro padrão da média ou simplesmente o erro padrão é dado por $\sigma/\sqrt{n}$.
- O erro padrão representa o desvio padrão estimado para um conjunto de médias amostrais de amostras de (tamanho igual a n) de uma população com variância σ^2.