Princípios de Bioestatística
Variável Aleatória Normal
Faixas de Referência

Enrico A. Colosimo/UFMG
http://www.est.ufmg.br/~enricoc/

Depto. Estatística - ICEx - UFMG

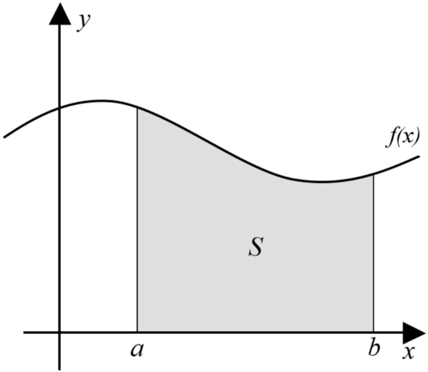
Variáveis Aleatórias Contínuas

- A probabilidade, no caso contínuo, é caracterizada a partir de uma função positiva denominada densidade de probabilidade. *'A densidade não é uma probabilidade, mas uma função matemática'.*

- Definição: f(X)é uma função contínua de probabilidade ou função densidade de probabilidade de uma variável aleatória contínua X, se:

    - $f(x) \geq 0, \forall x \epsilon \{-\infty; +\infty\}$ 0
    - A área definida por $f(x)$ é igual a 1: $\int_{-\infty}^{+\infty} f(x) = 1$
    - O cálculo de probabilidade é definido como: $P(a \leq X \leq b) = \int_a^b f(x)dx$ onde $a < b$

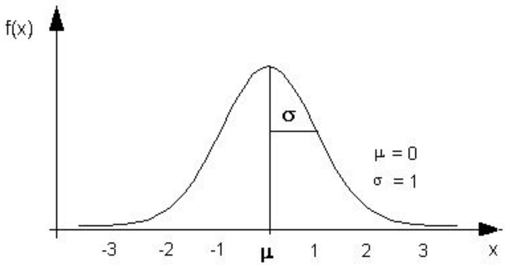
Cálculo de Probabilidade



- A probabilidade é definida como a área entre os pontos a e b

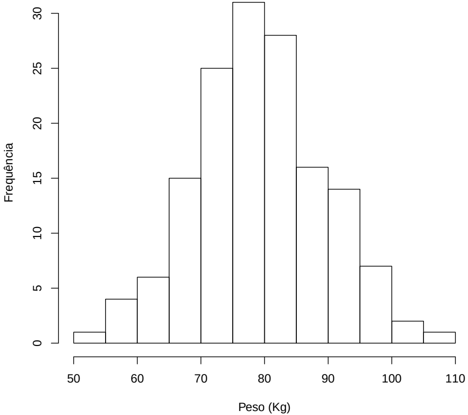
Distribuição Normal ou Curva de Gauss

- Definida por dois parâmetros: a média (μ) e o desvio padrão (σ).
- Simétrica em torno da média.
- Largura (amplitude) da curva determinada pelo desvio padrão.

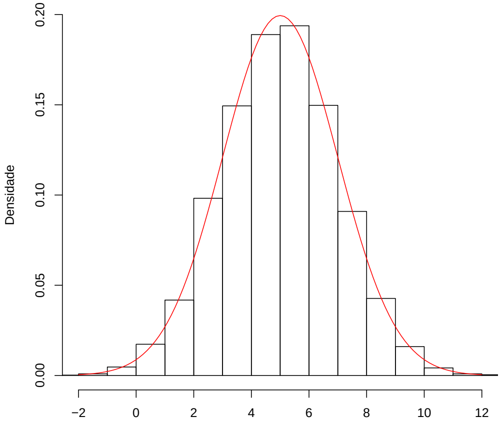


Histograma de Dados Amostrais

Exemplo: Histograma de uma amostra de pesos de 150 homens de uma certa população.



Histograma Suavizado

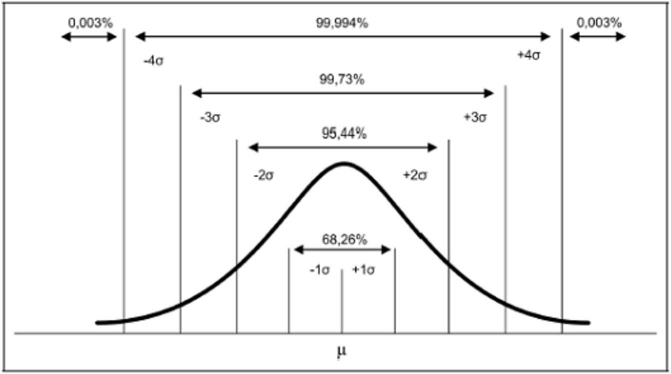


- A área sob a curva é igual à soma das áreas dos retângulos do histograma.
- Isto é, a área sob a curva compreende 100% dos dados.

Exemplos

- Estatura de adultos (segundo sexo)
- Peso ao Nascimento (segundo sexo)
- Comprimento da raiz do dente
- Perímetro cefálico (segundo raça)
- Faixas de referência
- Nota de Prova
- Etc, etc,...

Variável Normal com média μ e desvio-padrão σ



Função Matemática da Distribuição Normal

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

Modelo Gaussiano:

- Flexível $\rightarrow$ Modela vários fenômenos.
- Mínimo $\rightarrow$ Economia de parâmetros.
- Simétrico em torno da origem $\rightarrow$ Facilidade de leitura e interpretação.

Como calcular probabilidades?

$$P(a \leq X \leq b) = \int_a^b f(x)dx \text{ para } a < b$$

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

Felizmente, existe uma tabela pronta para ser consultada. Ou, de forma alternativa, podemos utilizar um software estatístico.

Distribuição Normal Padrão: $Z \sim N(\mu = 0, \sigma = 1)$

Seja X um variável Normal, com média μ e desvio padrão σ.

A variável Normal padronizada Z:

$$\frac{X - \mu}{\sigma} = Z \sim N(\mu = 0, \sigma = 1)$$

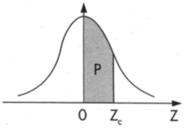
A transformação inversa é:

$$X = Z \cdot \sigma + \mu$$

Leitura da Tabela

Coluna 1: Inteiro e a primeira decimal
Cabeçalho das colunas: Segunda decimal
Corpo da Tabela: $P(0 \leq Z \leq Z_c) = \int_0^{Z_c} f(z)dz$

Tabela III — Distribuição Normal Padrão
$Z \sim N(0, 1)$
Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z < Z_c)$



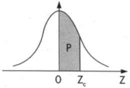
parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c										parte inteira e primeira decimal de Z_c
	0	1	2	3	4	5	6	7	8	9	
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1

Tabela

Tabela III — Distribuição Normal Padrão

$Z \sim N(0, 1)$

Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z < Z_c)$



parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c: 0	1	2	3	4	5	6	7	8	9	parte inteira e primeira decimal de Z_c
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	2,4
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	2,5
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	2,6
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736	2,7
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807	2,8
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861	2,9
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900	3,0
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929	3,1
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950	3,2

Exemplo 1

- X: pressão sanguínea diastólica (PD em mmHg);
- X tem Distribuição Normal
- Média $\mu = 77mmHg$
- Desvio Padrão $\sigma = 11,6mmHg$

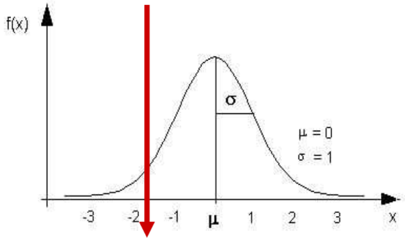
Probabilidade da PD ser inferior a 60 mmHg

Temos que:

- $60 - \mu = 60 \; - \; 77 \; = \; -17$
- $\frac{60 \; - \; 77}{\sigma} = -17 \; / \; 11,6 \; = \; -1,46$

Então,

- $Pr(Z < -1,46) \; = \; 0,0721$



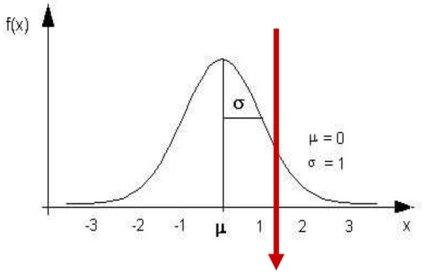
Probabilidade da PD ser superior a 90 mmHg

Temos que:

- $90 \ - \ 77 \ = \ 13$
- $13 \ / \ 11,6 \ = \ 1,12$

Então,

- $Pr(Z > 1,12) \ = \ 0,1314$



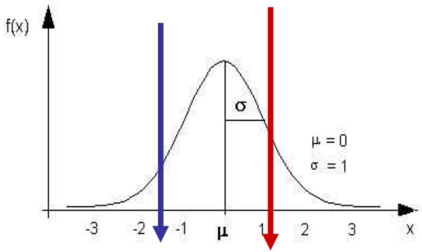
Probabilidade da PD ser superior a 60 mmHg e inferior a 90 mmHg

Temos que:

- $Pr(X > 90) = 0,8686$
- $Pr(X < 60) = 0,0721$

Então,

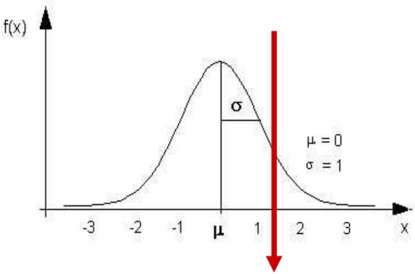
- $Pr(60 < X < 90) = Pr(-1,46 < Z < 1,12) =$
  $Pr(-1,46 < Z < 0) + P(0 < Z < 1,12) =$
  . $0,4278 + 0,3686 = 0,7964$



Qual é o valor da PD que deixa 90% dos indivíduos abaixo dela?

Temos que:

- Buscar o valor de Z na tabela (z=1,28)
- Fazer a transformação inversa
- $x = 1,28 * 11,6 \, + \, 77 = 92\text{mmHg}$



Exemplo 2

- X: valor de hemoglobina medido no sangue
- Distribuição Normal
- Média $\mu = 15\ g/100ml$
- Desvio Padrão $\sigma = 1,5\ g/100ml$

Proporção de indivíduos com hemoglobina maior que 18 g/100ml

Temos que:

- $18 \; - \; 15 \; = \; 3$
- $3 \; / \; 1,5 \; = \; 2$

Então,

- $Pr(X > 18) = Pr(Z > 2) \; = \; 0,0228$

Proporção de indivíduos com hemoglobina menor que 15 mg/100ml

Temos que:

- $15 - 15 = 0$

Então,

- $Pr(X < 15) = Pr(Z < 0) = 0,5$

Valor de hemoglobina que deixa 5% dos indivíduos acima dele

Temos que:

- $Pr(X > x_s) = 0,05$
- $\frac{x_s - 15}{1,5} = 1,645$
- $x_s = 17,46$
- 5% dos indivíduos da população tem hemoglobina superior a 17,5 mg/100ml

Faixas de referência

- Permite caracterizar o que é típico em uma determinada população.
- É empregado largamente em Ciências da Saúde, por exemplo, nos resultados de exames de laboratório.
- Esta metodologia tem outras aplicações, tais como a determinação de níveis toleráveis de barulho ou a caracterização dos níveis de poluição em uma região.

Faixas de referência

- Uma faixa de referência usual considerando uma cobertura de aproximadamente $(1-\alpha)100$ % é dada por:

$$[\mu \ - \ z_{\alpha/2} * \sigma \ ; \ \mu \ + \ z_{\alpha/2} * \sigma]$$

- Assumimos que a distribuição da quantidade de interesse é normal com média μ e desvio-padrão σ conhecidos.
- Por exemplo,
    - cobertura de 95% é dado por $[\mu \ - \ 1,96 * \sigma \ ; \ \mu \ + \ 1,96 * \sigma]$;
    - cobertura de 99% é dado por $[\mu \ - \ 2,57 * \sigma \ ; \ \mu \ + \ 2,57 * \sigma]$
    - cobertura sueprior de 99% é dado por $[\mu \ + \ 2,33 * \sigma]$

Faixas de referência - Exemplo

- X: teor de Gordura Fecal
- Distribuição Normal
- Média $\mu = 2,30\%$
- Desvio Padrão $\sigma = 0,87\%$

Faixas de referência - Exemplo

- Faixa de referência com aproximadamente 95% de cobertura
- $[\mu \; - \; 1,96 * \sigma \; ; \; \mu \; + \; 1,96 * \sigma]$
- $[2,30 \; - \; 1,96 * 0,87 \; ; \; 2,30 \; + \; 1,96 * 0,87]$
- $[0,56 \; ; \; 4,05]\%$

Intervalos de Referência

CE Update

Received 12.26.05 | Revisions Received 1.5.06 | Accepted 1.6.06

Statistical Methods for Establishing and Validating Reference Intervals

Roger L. Bertholf, PhD
(University of Florida Health Science Center/Jacksonville, Jacksonville, FL)
DOI: 10.1309/CBMHPRFNLU1XA4XV

Establishing Reference Intervals

Reference intervals customarily represent the central 95% of values obtained from the reference population. Consequently, 2.5% of "normal" individuals will exceed the reference range, and 2.5% will be below it. It is tempting to assume that normal values for clinical laboratory measurements conform to a Gaussian distribution, in which the central 95% of the area under the probability distribution curve corresponds to the population mean (μ) $\pm$ 1.96 standard deviations (usually rounded to 2 SD, or 2σ). However, this approach is often misguided, since the concentrations of various biochemicals in the body rarely follow a Gaussian distribution, due to physiological factors that influence the concentration in a unidirectional manner; intra-individual variations are not strictly random. Statistical approaches that are based on a predictable distribution of data, such as the Gaussian (or "Normal") distribution, are called "parametric," since they make certain assumptions about the data derived from the population. Non-parametric methods make no assumptions about how the data are distributed, and provide ways to analyze and compare data sets that have unknown or unpredictable distributions.

Faixas de referência

Usualmente são obtidas de duas formas.

- Distribuição Normal/Paramétrico
  - Medida de interesse tem distribuição normal
  - Prática de ser obtida e necessita um tamanho de amostra menor.
- Percentis/Não-Paramétrico
  - Flexível: pois pode ser utilizada para qualquer medida de interesse.
  - Necessita de muita informação, tamanho de amostra grande.