

# Introdução à Bioestatística

## Inferência e Intervalo de Confiança

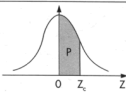
Enrico A. Colosimo/UFMG

<http://www.est.ufmg.br/~enricoc/>

Depto. Estatística - ICEx - UFMG

# Tabela Normal Padrão

**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
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
3,3	49952	49953	49955	49957	49958	49959	49960	49961	49962	49963	3,3

- 1 Observe que supomos conhecidos:
  - probabilidade de nascer um menino (0.5);
  - média e desvio-padrão do teor de gordura;
  - média e desvio-padrão da pressão diastólica, etc.
- 2 No mundo real, usualmente, estas quantidades não são conhecidas.
- 3 Na verdade, o nosso objetivo, é exatamente buscar informações sobre estas quantidades a partir de uma amostra da população.
- 4 Lembre: (primeiras aulas)
  - comparar cirurgia e fisioterapia para pacientes com incontinência urinária.
  - prevalência de depressão em idosos (Profa. Marisa).
  - preferência do eleitorado por um candidato X, etc.

- 1 Utilizar Método Estatístico + Amostra para responder à pergunta de interesse.
- 2 Inferência Estatística consiste, essencialmente, em:
  - Estimativa pontual e intervalar (intervalo de confiança).
  - Teste de hipóteses.
  - Previsão.

## Resultado Fundamental 1: Distribuição de $\bar{X}$

Suponha uma amostra aleatória de tamanho  $n$  retirada de uma **população normal** com média  $\mu$  e varância  $\sigma^2$ . Então

$$\bar{X} \sim N(\mu, \sigma/\sqrt{n})$$

ou

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim Z$$

.

## Resultado Fundamental 2: Teorema Central do Limite

Suponha uma amostra aleatória de tamanho  $n$  retirada de uma **população qualquer** com média  $\mu$  e variância  $\sigma^2$ . Representando tal amostra por  $n$  variáveis aleatórias independentes  $(X_1, \dots, X_n)$  e, denotando sua média por  $\bar{X}$ , temos que:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \rightarrow Z$$

quando  $n \rightarrow \infty$  e com  $Z \sim N(0, 1)$ .

## Teorema Central do Limite

- Em palavras, o teorema garante que para  $n$  grande a distribuição da média amostral, devidamente padronizada, se comporta segundo um modelo Normal com média 0 e variância 1.
- Pelo teorema temos que quanto maior o tamanho da amostra, melhor é a aproximação.
- Estudos envolvendo simulações, mostram que, em muitos casos, valores de  $n$  ao redor de 30 fornecem aproximações bastante boas para as aplicações práticas.

# Simulação

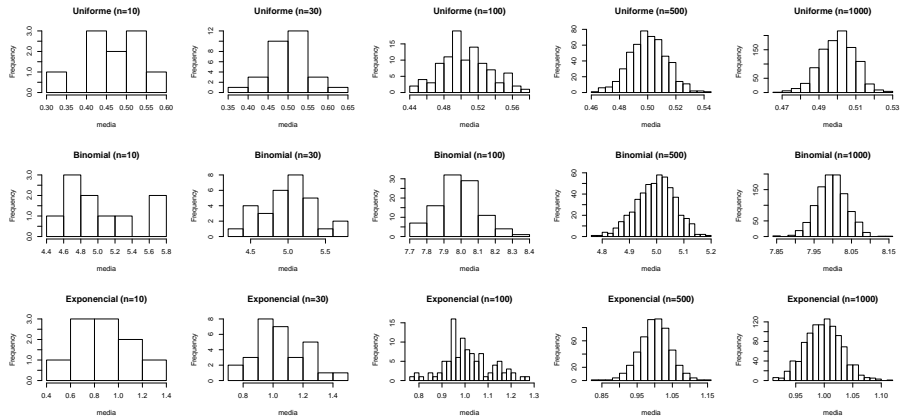


Figura: Efeito do tamanho da amostra sobre a distribuição de  $\bar{X}$



## Exemplo

- Suponha que para crianças nascidas com peso abaixo de  $750g$ , o nível de bilirrubina sérico tem distribuição Normal com média  $8,5mg/dl$  e desvio padrão  $3,5mg/dl$ .
- Calcule a probabilidade que a média amostra  $\bar{X}$ , para uma amostra de 16 crianças:

## Seja menor que 8 mg/dl

Temos que:

- $\bar{X} \sim N(8,5; 3,5/4)$
- $8 - 8,5 = -0,5$
- $3,5/4 = 0,875$
- $-0,5 / 0,875 = -0,5714$

Então,

- $Pr(\bar{X} < 8) = Pr(Z < -0,5714) = Pr(Z > 0,57) = 0,5 - 0,2157 = 0,28$

## Seja maior que 8,2 mg/dl

Temos que:

- $8,2 - 8,5 = -0,3$
- $3,5/4 = 0,875$
- $-0,3 / 0,875 = -0,3428$

Então,

- $Pr(\bar{X} > 8,2) = Pr(Z > -0,3428) = P(0 < Z < 0,34) + 0,5 = 0,6330$

E se a distribuição de bilirrubina não tivesse distribuição normal?

## Intervalo de Confiança para a Média $\mu$

Em uma certa população sabemos que a pressão diastólica de seus habitantes tem uma distribuição normal com média desconhecida  $\mu$  e desvio-padrão  $\sigma = 10,2$  mmHg. Forneça uma estimativa pontual e intervalar (95% de confiança) para  $\mu$  baseado em uma amostra de tamanho 100 em que foi obtido  $\bar{x} = 82,4$  mmHg.

- Estimativa pontual: 82,4 mmHg
- Estimativa intervalar:  
 $[\bar{x} - 1,96 * \sigma/10 ; \bar{x} + 1,96 * \sigma/10]$   
 $[82,4 - 1,96 * 1,02 ; 82,4 + 1,96 * 1,02]$   
 $[80,4 ; 84,4]$  mmHg

## Intervalo de Confiança para a Média $\mu$

- 1 Como estimar  $\mu$  se a distribuição não for Normal?

Resposta:  $n = 100$  deve ser suficientemente grande para  $\bar{X}$  ter uma distribuição aproximadamente normal pelo Teorema Central do Limite. Desta forma, os resultados anteriores continuam válidos.

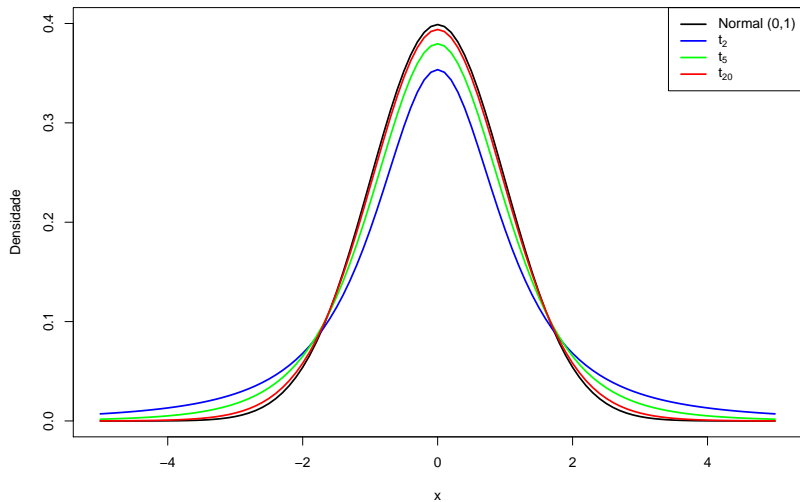
- 2 Como estimar  $\mu$  se o desvio-padrão  $\sigma$  for desconhecido e  $n$  não for suficientemente grande para usarmos o TCL?

## Resultado Fundamental 3: Distribuição T- Student

- Estimamos a média amostral  $\bar{x}$  e o desvio-padrão amostral  $s$ ;
- E substituímos  $\sigma$  por  $s$  e a distribuição normal pela distribuição t-Student com  $n - 1$  graus de liberdade.
- Ou seja,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

## Distribuição Normal versus t



# Tabela T- Student

Percentile Values ( $t_p$ )  
for  
Student's  $t$  Distribution  
with  $\nu$  Degrees of Freedom  
(shaded area =  $p$ )



$\nu$	$t_{.995}$	$t_{.99}$	$t_{.975}$	$t_{.95}$	$t_{.90}$	$t_{.80}$	$t_{.75}$	$t_{.70}$	$t_{.60}$	$t_{.55}$
1	63.66	31.82	12.71	6.31	3.08	1.376	1.000	.727	.325	.151
2	9.92	6.96	4.30	2.92	1.89	1.061	.816	.617	.289	.142
3	5.84	4.54	3.18	2.35	1.64	.978	.765	.584	.277	.137
4	4.60	3.75	2.78	2.13	1.53	.941	.741	.569	.271	.134
5	4.03	3.36	2.57	2.02	1.48	.920	.727	.559	.267	.132
6	3.71	3.14	2.45	1.94	1.44	.906	.718	.553	.265	.131
7	3.50	3.00	2.36	1.90	1.42	.896	.711	.549	.263	.130
8	3.36	2.90	2.31	1.86	1.40	.889	.706	.546	.262	.130
9	3.25	2.82	2.26	1.83	1.38	.883	.703	.543	.261	.129
10	3.17	2.76	2.23	1.81	1.37	.879	.700	.542	.260	.129
11	3.11	2.72	2.20	1.80	1.36	.876	.697	.540	.260	.129
12	3.06	2.68	2.18	1.78	1.36	.873	.695	.539	.259	.128
13	3.01	2.65	2.16	1.77	1.35	.870	.694	.538	.259	.128
14	2.98	2.62	2.14	1.76	1.34	.868	.692	.537	.258	.128
15	2.95	2.60	2.13	1.75	1.34	.866	.691	.536	.258	.128
16	2.92	2.58	2.12	1.75	1.34	.865	.690	.535	.258	.128
17	2.90	2.57	2.11	1.74	1.33	.863	.689	.534	.257	.128
18	2.88	2.55	2.10	1.73	1.33	.862	.688	.534	.257	.127
19	2.86	2.54	2.09	1.73	1.33	.861	.688	.533	.257	.127
20	2.84	2.53	2.09	1.72	1.32	.860	.687	.533	.257	.127
21	2.83	2.52	2.08	1.72	1.32	.859	.686	.532	.257	.127
22	2.82	2.51	2.07	1.72	1.32	.858	.686	.532	.256	.127
23	2.81	2.50	2.07	1.71	1.32	.858	.685	.532	.256	.127
24	2.80	2.49	2.06	1.71	1.32	.857	.685	.531	.256	.127
25	2.79	2.48	2.06	1.71	1.32	.856	.684	.531	.256	.127
26	2.78	2.48	2.06	1.71	1.32	.856	.684	.531	.256	.127
27	2.77	2.47	2.05	1.70	1.31	.855	.684	.531	.256	.127
28	2.76	2.47	2.05	1.70	1.31	.855	.683	.530	.256	.127
29	2.76	2.46	2.04	1.70	1.31	.854	.683	.530	.256	.127
30	2.75	2.46	2.04	1.70	1.31	.854	.683	.530	.256	.127
40	2.70	2.42	2.02	1.68	1.30	.851	.681	.529	.255	.126
60	2.66	2.39	2.00	1.67	1.30	.848	.679	.527	.254	.126
120	2.62	2.36	1.98	1.66	1.29	.845	.677	.526	.254	.126
$\infty$	2.58	2.33	1.96	1.645	1.28	.842	.674	.524	.253	.126

Source: R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research* (5th edition), Table III, Oliver and Boyd Ltd., Edinburgh, by permission of the authors and publishers.



## Intervalo de Confiança para a Média $\mu$ com $\sigma$ desconhecido.

O crescimento de bebês durante o primeiro mês de vida pode ser modelado pela distribuição Normal.

Uma amostra de 10 recém nascidos na região foram sorteados e sua altura acompanhada, fornecendo as seguintes medidas de crescimento em centímetros: 5,03 ; 5,02 ; 4,95 ; 4,96 ; 5,01 ; 4,97 ; 4,9 ; 4,91 ; 4,9 ; 4,93 ( $\bar{x} = 4,96$ ;  $s = 0,049$ )

Forneça uma estimativa pontual e intervalar (95% de confiança) para  $\mu$

## Intervalo de Confiança para a Média $\mu$ com $\sigma$ desconhecido.

- Estimativa pontual: 4,96 cm

- Estimativa intervalar:

$$[\bar{x} - t_{9;0,975} * s/\sqrt{10} ; \bar{x} + t_{9;0,975} * s/\sqrt{10}]$$

$$[4,96 - 2,26 * 0,049/3,16 ; 4,96 + 2,26 * 0,049/3,16]$$

$$[4,96 - 2,26 * 0,015 ; 4,96 + 2,26 * 0,015]$$

$$[4,96 - 0,035 ; 4,96 + 0,035]$$

$$[4,92 ; 4,99] \text{ cm}$$