9. INTERVALOS DE PREDIÇÃO E INTERVALOS DE TOLERÂNCIA

2010

Intervalo de confiança

$X_1, \cdots X_n$ é uma amostra aleatória de tamanho n de uma população normal com média μ e variância σ^2 (ambas desconhecidas). A média amostral $\bar{X}$ tem distribuição normal com média μ e variância σ^2 / n:

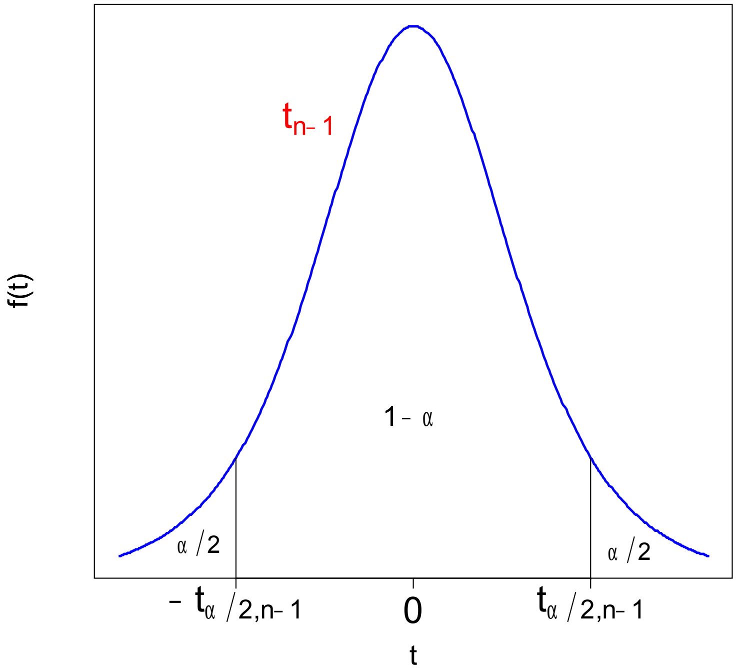
$$Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0,1).$$

$T = \dfrac{\sqrt{n}(\bar{X} - \mu)}{s} \sim t_{n-1},$: distribuição t de Student com $n - 1$ g.l.

sendo que *s* é o desvio padrão amostral.

Escolhendo um coeficiente de confiança $(1-\alpha)$ pode-se determinar $t_{\alpha/2,n-1}$ (consultando a Tábua III) tal que

$$P(-t_{\alpha/2,\, n-1} \le T \le t_{\alpha/2,\, n-1}) = 1 - \alpha.$$



Intervalo de confiança

Substituindo T e isolando μ obtemos

$$\mathrm{P}\left(\overline{X} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \right) = 1-\alpha\ .$$

Logo, um intervalo de confiança (IC) de $100 \times (1-\alpha)\%$ para a média populacional μ é dado por

$$[L;U] = \left[\overline{X} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}; \overline{X} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \right] = \left[\overline{X} - E; \overline{X} + E\right],$$

sendo que $E = t_{\alpha/2,n-1}\dfrac{s}{\sqrt{n}}$ é o erro máximo.

Outros problemas:

1. Apresentar um intervalo de possíveis valores para a observação X_{n+1} (nova observação ou observação futura).

2. Apresentar um intervalo que contenha uma certa proporção de valores de X.

9.2. Intervalos de predição

X_{n+1} é uma nova observação e $X_{n+1} \sim N(\mu\ \sigma^2)$.

Predição: $\overline{X}$. Erro de predição: $X_{n+1} - \overline{X}$.

Resultado: $X_{n+1} - \overline{X} \sim \mathrm{N}\left(0, \sigma^2 + \dfrac{\sigma^2}{n}\right)$. Padronização: $Z = \dfrac{X_{n+1} - \overline{X}}{\sigma\sqrt{1 + \dfrac{1}{n}}} \sim \mathrm{N}(0,1)$.

Substituindo σ por s: $T = \dfrac{X_{n+1} - \overline{X}}{s\sqrt{1 + \dfrac{1}{n}}} \sim t_{n-1}$.

Na Tábua III obtemos $t_{\alpha/2,\, n-1}$ tal que $P(-t_{\alpha/2,\, n-1} \leq T \leq t_{\alpha/2,\, n-1}) = 1 - \alpha$.

Substituindo T e isolando X_{n+1} obtemos

$$\mathrm{P}\left(\overline{X} - t_{\alpha/2, n-1} \times s \times \sqrt{1 + \frac{1}{n}} \leq X_{n+1} \leq \overline{X} + t_{\alpha/2, n-1} \times s \times \sqrt{1 + \frac{1}{n}}\right) = 1 - \alpha\,.$$

Um intervalo de predição de $100 \times (1-\alpha)\%$ para X_{n+1} é dado por

$$[L;U] = [\overline{X} - E; \overline{X} + E], \quad \text{em que } E = t_{\alpha/2, n-1} \times s \times \sqrt{1 + \frac{1}{n}}.$$

9.3. Intervalos de tolerância

Se X ~ N(μ σ²) vimos no Cap, 5 que

$$P(\mu - \sigma \le X \le \mu + \sigma) \quad = 0{,}6896,$$
$$P(\mu - 2\sigma \le X \le \mu + 2\sigma) = 0{,}9546 \text{ e}$$
$$P(\mu - 3\sigma \le X \le \mu + 3\sigma) = 0{,}9973.$$

O intervalo [μ – 2 σ; μ + 2 σ] cobre 95,46% dos valores de X.

Entretanto, μ e σ são (geralmente) desconhecidos. Substituindo pelos estimadores obtemos $[\bar{X} - 2s; \bar{X} + 2s]$, cuja cobertura não é 95,46%.

Para uma dada cobertura, o intervalo é $[\bar{X} - k \times s; \bar{X} + k \times s]$.

Como este intervalo é aleatório, temos mais uma fonte de incerteza, de modo que é necessário apresentar um coeficiente de confiança para o intervalo.

Uma vez escolhidas a cobertura e o coeficiente de confiança, o valor de *k* é obtido da tabela na lâmina 6.
A cobertura e o coeficiente de confiança devem ser “altos” (90%, 95%, ou 99%, por exemplo).

Valores de k

Tamanho da amostra (n)	Coeficiente de confiança 0.90			0.95			0.99		
	Probabilidade de cobertura 0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
2	15.978	18.800	24.167	32.019	37.674	48.430	160.193	188.491	242.300
3	5.847	6.919	8.974	8.380	9.916	12.861	18.930	22.401	29.055
4	4.166	4.943	6.440	5.369	6.370	8.299	9.398	11.150	14.527
5	3.949	4.152	5.423	4.275	5.079	6.634	6.612	7.855	10.260
6	3.131	3.723	4.870	3.712	4.414	5.775	5.337	6.345	8.301
7	2.902	3.452	4.521	3.369	4.007	5.248	4.613	5.488	7.187
8	2.743	3.264	4.278	3.136	3.732	4.891	4.147	4.936	6.468
9	2.626	3.125	4.098	2.967	3.532	4.631	3.822	4.550	5.966
10	2.535	3.018	3.959	2.839	3.379	4.433	3.582	4.265	5.594
11	2.463	2.933	3.849	2.737	3.259	4.277	3.397	4.045	5.308
12	2.404	2.863	3.758	2.655	3.162	4.150	3.250	3.870	5.079
13	2.355	2.805	3.682	2.587	3.081	4.044	3.130	3.727	4.893
14	2.314	2.756	3.618	2.529	3.012	3.955	3.029	3.608	4.737
15	2.278	2.713	3.562	2.480	2.954	3.878	2.945	3.507	4.605
16	2.246	2.676	3.514	2.437	2.903	3.812	2.872	3.421	4.492
17	2.219	2.643	3.471	2.400	2.858	3.754	2.808	3.345	4.393
18	2.194	2.614	3.433	2.366	2.819	3.702	2.753	3.279	4.307
19	2.172	2.588	3.399	2.337	2.784	3.656	2.703	3.221	4.230
20	2.152	2.564	3.368	2.310	2.752	3.615	2.659	3.168	4.161
21	2.135	2.543	3.340	2.286	2.723	3.577	2.620	3.121	4.100
22	2.118	2.524	3.315	2.264	2.697	3.543	2.584	3.078	4.044
23	2.103	2.506	3.292	2.244	2.673	3.512	2.551	3.040	3.993
24	2.089	2.489	3.270	2.225	2.651	3.483	2.522	3.004	3.947
25	2.077	2.474	3.251	2.208	2.631	3.457	2.494	2.972	3.904
30	2.025	2.413	3.170	2.140	2.529	3.350	2.385	2.841	3.733
40	1.959	2.334	3.066	2.052	2.445	3.213	2.247	2.677	3.518
50	1.916	2.284	3.001	1.996	2.379	3.126	2.162	2.576	3.385
60	1.887	2.248	2.955	1.958	2.333	3.066	2.103	2.506	3.293
70	1.865	2.222	2.920	1.929	2.299	3.021	2.060	2.454	3.225
80	1.848	2.202	2.894	1.907	2.272	2.986	2.026	2.414	3.173
90	1.834	2.185	2.872	1.889	2.251	2.958	1.999	2.382	3.130
100	1.822	2.172	2.854	1.874	2.233	2.934	1.977	2.355	3.096

Exemplo. Em uma amostra com n = 20 observações, para uma cobertura de 90% e uma confiança de 95% encontramos k = 2,310.

Intervalo de tolerância :

$$[\overline{X} - 2{,}310 \times s; \overline{X} + 2{,}310 \times s].$$

Utilizando este intervalo em muitas amostras com n = 20 teremos uma proporção de 95% dos intervalos contendo 90% de todos os valores de X.

Exemplo

Uma máquina produz bastões cilíndricos metálicos. Uma amostra aleatória de 15 bastões foi coletada. Os diâmetros (em mm) foram medidos resultando em

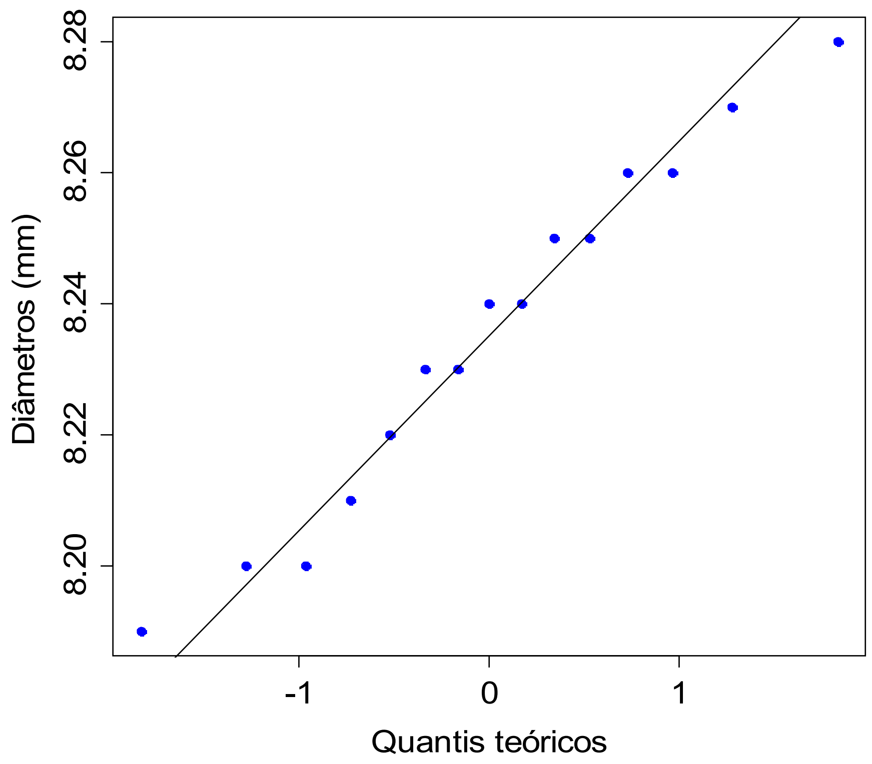
`8,24  8,23  8,20  8,21  8,20  8,28  8,22  8,26  8,27  8,25`
`8,19  8,25  8,26  8,23  8,24.`

Utilizando os dados coletados, responda às seguintes questões:

(a) O que pode ser afirmado sobre o diâmetro de um bastão adicional a ser coletado?

(b) É de interesse acompanhar o comportamento do diâmetro médio dos bastões produzidos.

(c) Que informação você comunicaria a um possível comprador dos bastões?



Solução. X representa a variável diâmetro dos bastões, em mm. Supomos que $X \sim N(\mu\ \ \sigma^2)$. O gráfico de quantis da figura ao lado não contradiz esta suposição.

Exemplo

Foram coletadas n = 15 observações. Calculamos

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 8{,}235 \text{ mm} \quad \text{e} \quad s = \left\{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right\}^{1/2} = 0{,}0272 \text{ mm}.$$

Na letra (a) a quantidade de interesse é X_{16}. Apresentaremos um intervalo de predição de 95% para X_{16} .

Na Tábua III com n – 1 = 14 graus de liberdade e p / 2 = (1 – 0,95) / 2 = 0,05 / 2 (ou seja, p = 5%), obtemos $t_{\alpha/2,\ n-1}$ = 2,145.

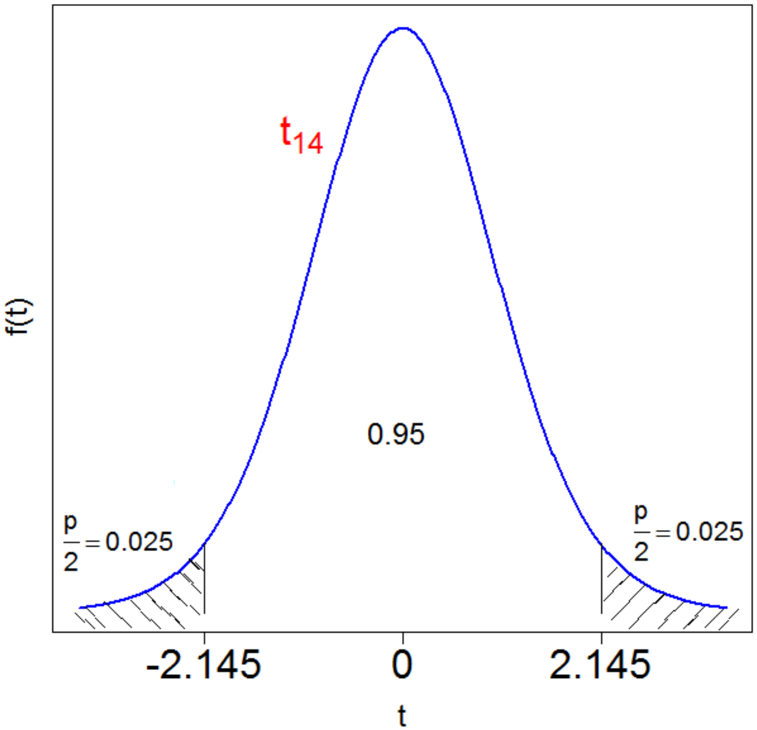
Calculamos o erro máximo (lâmina 4)

$$E = t_{\alpha/2,n-1} \times s \times \sqrt{1 + \frac{1}{n}} = 2{,}145 \times 0{,}0272 \times \sqrt{1 + \frac{1}{15}}$$

$$= 0{,}0603 \text{ mm}.$$

Um intervalo de predição de 95% para X_{16} é dado por

$$[\bar{X} - E; \bar{X} + E] = [8{,}175;\ 8{,}296], \text{ em mm}.$$



Exemplo

Na letra (b) a quantidade de interesse é o diâmetro médio (μ) dos bastões. Assim, Apresentaremos um intervalo de confiança (IC) de 95% para μ.

Já vimos que $t_{\alpha/2,\, n-1} = 2{,}145$. Calculamos o erro máximo (lâmina 3)

$$E = t_{\alpha/2, n-1} \times s \times \sqrt{\frac{1}{n}} = 2{,}145 \times 0{,}0272 \times \sqrt{\frac{1}{15}}$$

$$= 0{,}0151 \text{ mm.}$$

Obs. Como esperado, o erro máximo do IC é menor do que o erro máximo do intervalo de predição (lâmina 8).

Um IC de 95% para μ é dado por $[\overline{X} - E; \overline{X} + E] = [8{,}220;\ 8{,}250]$, em mm.

Na letra (c), para um possível comprador pode ser útil conhecer um intervalo que contenha uma proporção alta dos valores do diâmetro. Apresentaremos um intervalo de tolerância de 95% para X com uma cobertura de 99%.

Consultando a tabela da lâmina 6 com n = 15, confiança = 0,95 e cobertura = 0,99, encontramos $k = 3{,}878$. Um intervalo de tolerância para o diâmetro dos bastões é dado por (lâmina 5)

$$[\overline{X} - k \times s; \overline{X} + k \times s] = [8{,}130;\ 8{,}341], \text{ em mm.}$$

Obs. Se a distribuição de X não for normal, os intervalos deste capítulo são soluções aproximadas.