CC0288 - Inferência Estatística I

Aula - 08/05/2023.

Prof. Maurício

Esta lista é composta pelos exercícios do Capítulo V do Navidi, Probabilidade e Estatística para Ciências Exatas. São exercícios bem práticos. Vocês gostarão de resolvê-los. Vamos colocar os exercícios da seção 5.5.

Ele trata da construção dos Intervalos de Predição e Intervalos de Tolerância.

Um intervalo de confiança para um parâmetro como a média populacional é um interva lo que pode conter o valor real do parâmetro.

Em contraste, os intervalos de predição e tolerância estão relacionados com a própria população e com os valores que podem ser amostrados dela no futuro.

Os intervalos de predição e tolerância são úteis apenas quando a forma da população é conhecida. Os métodos que apresentamos aqui, que são os mais usados, são válidos apenas quando se sabe que a população tem distribuição normal.

Intervalos de predição;

Um intervalo de predição é um intervalo que pode conter o valor de um item que será amostrado de uma população em um momento futuro. Em outras palavras, "prevemos"que um valor que ainda será amostrado da população se encontra dentro de um intervalo de predição.

Vamos ilustrar isso com um exemplo:

Exemplo 1: Considere que o teor de silício (X) medida em porcentagem foi medido para cinco vigas de aço e que a média amostral é $\bar{x} = 0,26$ com um desvio padrão $s = 0,05$.

Em algum momento futuro observaremos o teor de silício (Y) de alguma outra viga e queremos construir um intervalo que contenha o valor de Y com probabilidade 0,95.

Para ver como isso é feito
suponha que

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

e que

$$Y \sim N(\mu, \sigma^2).$$

Na realidade retiramos uma amostra aleatória simples de X, $X_1, X_2, \ldots, X_n$ e temos que

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Considere

$$X_{n+1} = Y \sim N(\mu, \sigma^2),$$

de forma que $X_{n+1} = Y$ e $\bar{X}$ sejam independentes.

Logo

$$Y - \bar{X} \sim N\left(0, \sigma^2 + \frac{\sigma^2}{n}\right).$$

$$Y - \bar{X} \sim N\left(0, \sigma^2\left[1 + \frac{1}{n}\right]\right).$$

Padronizando temos:

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

Sabemos que

$$V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

Assim

$$T = \frac{Y - \bar{X}}{S\sqrt{1 + \frac{1}{n}}} \sim t(n-1).$$

Considere que

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

$$P\left(-t_{n-1,\alpha/2} \leq \frac{Y - \bar{X}}{S\sqrt{1 + \frac{1}{n}}} \leq t_{n-1,\alpha/2}\right) = 1 - \alpha.$$

$$P\left(\bar{X} - t_{n-1,\alpha/2} \times S\sqrt{1 + \frac{1}{n}} \leq Y \leq \bar{X} + t_{n-1,\alpha/2} \times S\sqrt{1 + \frac{1}{n}}\right) = 1 - \alpha.$$

Vamos resumir:

Seja $X_1, X_2, \ldots, X_n$ um amostra de uma população normal . Seja $Y = X_{n+1}$ outro item a ser amostrado desta população cujo valor ainda não foi observado.

Um intervalo de predição ao nível $\gamma = 100(1 - \alpha)\%$ para $Y = X_{n+1}$ é:

$$\bar{X} \pm t_{n-1,\alpha/2} \times S\sqrt{1 + \frac{1}{n}}.$$

Note que a probabilidade de Y pertencer ao intervalo é $1 - \alpha$.

O intervalo de previsão é para um valor da variável e não para um parâmetro.

Vamos usar o R para terminar o exemplo:

```
> n=5;xb=0.26;s=0.05; gama=0.95;alfa=1-gama;alfa;alfa/2;1-alfa/2
[1] 0.05
[1] 0.025
[1] 0.975
>
> t_tab=qt(1-alfa/2,n-1);t_tab;round(t_tab,3)
```

```
[1] 2.776445
[1] 2.776
>
> e=t_tab*s*sqrt(1+1/n);e
[1] 0.1520722
>
> IP95=xb+c(-1,1)*e;IP95;round(IP95,2)
[1] 0.1079278 0.4120722
[1] 0.11 0.41
>
```

Intervalos de Predição Unilateral:

Seja $X_1, X_2, \ldots, X_n$ um amostra de uma população normal . Seja $Y = X_{n+1}$ outro item a ser amostrado desta população cujo valor ainda não foi observado.

Um limite de predição superior ao nível $\gamma = 100(1-\alpha)\%$ para $Y = X_{n+1}$ é:

$$\bar{X} + t_{n-1,\alpha} \times S\sqrt{1+\frac{1}{n}}.$$

Um limite de predição inferior ao nível $\gamma = 100(1-\alpha)\%$ para $Y = X_{n+1}$ é:

$$\bar{X} - t_{n-1,\alpha} \times S\sqrt{1+\frac{1}{n}}.$$

Intervalos de Tolerância para uma população Normal.

Um intervalos de tolerância é aquele que provavelmente contém uma proporção especificada da população.

O método que apresentamos agora, que é o mais usado, requer que a população seja Normal.

Para ilustrar a ideia, primeiro considere que temos uma população normal com média μ e desvio padrão σ conhecidos.

Se quisermos determinar um intervalo que contenha 90%, podemos assim fazê-lo

O intervalo

$$\mu \pm 1,6445\sigma,$$

contém 90% da população.

O intervalo

$$\mu \pm z_{1-\alpha/2}\sigma,$$

contém $100(1-\alpha)\%$ da população.

Na prática não conhecemos os valores de μ e σ. Em vez disso, temos uma amostra de tamanho n e estimamos μ com a média amostral $\bar{x}$ e σ pelo desvio padrão amostral s. Essa estimação tem duas consequências:

Primeiro, temos que fazer com que o intervalo seja mais extenso do que μ e σ conhecidos. Segundo, não podemos ter de 100% de que o intervalo contém a proporção desejada da população.

Portanto, para construir um intervalo de tolerância , temos que especificar a proporção $100(1-\beta)\%$ da população que queremos que o intervalo contenha, juntamente com um nível de confiança de $100(1-\alpha)\%$ de que o intervalo realmente contenha a proporção especificada.

Então é possível, usando métodos avançados, determinar um número $k_{n,\alpha,\beta}$ de modo que o intervalo

$$\bar{X} \pm k_{n,\alpha,\beta} \times s$$

contenha pelo menos $100(1-\beta)\%$ da população com confiança de $100(1-\alpha)\%$. Os valores de $k_{n,\alpha,\beta}$ estão apresentados na tabela a seguir:

Tabela A.4 Fatores de tolerância para a distribuição normal

Tamanho da amostra n	Nível de confiança de 95% — Percentual da população contida 90%	95%	99%	Nível de confiança de 99% — Percentual da população contida 90%	95%	99%
2	32,0187	37,6746	48,4296	160,1940	188,4915	242,3004
3	8,3795	9,9158	12,8613	18,9304	22,4009	29,0553
4	5,3692	6,3699	8,2993	9,3984	11,1501	14,5274
5	4,2749	5,0787	6,6338	6,6118	7,8550	10,2602
6	3,7123	4,4140	5,7746	5,3366	6,3453	8,3013
7	3,3686	4,0074	5,2481	4,6129	5,4877	7,1868
8	3,1358	3,7317	4,8907	4,1473	4,9355	6,4683
9	2,9670	3,5317	4,6310	3,8223	4,5499	5,9660
10	2,8385	3,3794	4,4330	3,5821	4,2647	5,5943
11	2,7372	3,2592	4,2766	3,3970	4,0449	5,3075
12	2,6550	3,1617	4,1496	3,2497	3,8700	5,0792
13	2,5868	3,0808	4,0441	3,1295	3,7271	4,8926
14	2,5292	3,0124	3,9549	3,0294	3,6081	4,7371
15	2,4799	2,9538	3,8785	2,9446	3,5073	4,6053
16	2,4371	2,9029	3,8121	2,8717	3,4207	4,4920
17	2,3995	2,8583	3,7538	2,8084	3,3453	4,3934
18	2,3662	2,8188	3,7022	2,7527	3,2792	4,3068
19	2,3366	2,7835	3,6560	2,7034	3,2205	4,2300
20	2,3099	2,7518	3,6146	2,6594	3,1681	4,1614
25	2,2083	2,6310	3,4565	2,4941	2,9715	3,9039
30	2,1398	2,5494	3,3497	2,3848	2,8414	3,7333
35	2,0899	2,4900	3,2719	2,3063	2,7479	3,6107
40	2,0516	2,4445	3,2122	2,2468	2,6770	3,5177
45	2,0212	2,4083	3,1647	2,1998	2,6211	3,4443
50	1,9964	2,3787	3,1259	2,1616	2,5756	3,3846
60	1,9578	2,3328	3,0657	2,1029	2,5057	3,2929
70	1,9291	2,2987	3,0208	2,0596	2,4541	3,2251
80	1,9068	2,2720	2,9859	2,0260	2,4141	3,1725
90	1,8887	2,2506	2,9577	1,9990	2,3819	3,1303
100	1,8738	2,2328	2,9343	1,9768	2,3555	3,0955
200	1,7981	2,1425	2,8158	1,8651	2,2224	2,9207
300	1,7670	2,1055	2,7671	1,8199	2,1685	2,8499
400	1,7492	2,0843	2,7392	1,7940	2,1377	2,8094
500	1,7373	2,0701	2,7206	1,7769	2,1173	2,7826
600	1,7287	2,0598	2,7071	1,7644	2,1024	2,7631
700	1,7220	2,0519	2,6967	1,7549	2,0911	2,7481
800	1,7167	2,0456	2,6884	1,7473	2,0820	2,7362
900	1,7124	2,0404	2,6816	1,7410	2,0746	2,7264
1000	1,7087	2,0361	2,6759	1,7358	2,0683	2,7182

Figura 1:

Exemplo 2: Os comprimentos dos parafusos fabricados por meio de determinado processo são conhecidos como tendo uma distribuição normal. Em uma amostra de 30 parafusos o comprimento médio foi de 10,25 cm , com desvio padrão 0,20 cm. Determinar um intervalo de tolerância que inclua 90% com confiança de 95%.

Solução:

Temos que

$$n = 30 \quad ; \quad \bar{x} = 10,25 \quad ; \beta = 1 - 0,90 = 0,10 \quad ; \quad \alpha = 1 - 0,95 = 00,05.$$

Olhando a tabela temos:

$$k_{n,\alpha,\beta} = k_{3,0,05,0,10} = 2,1398.$$

O intervalo de tolerância é dado por:

$$\bar{X} \pm k_{n,\alpha,\beta} \times s = 10,25 \pm 2,1398 \times 0,20 = 10 \pm .$$

```
>
> xb=10.25
>
> IT9590=xb+c(-1,1)*e;IT9590
[1]  9.82204 10.67796
> round(IT9590,2)
[1]  9.82 10.68
>
```

Agora faça os exercícios:

1. (Seção 5.5- Exercício 1.) Uma amostra de 25 resistores, com valor nominal de 100 Ω tem uma resistência média de 101,4 Ω e um desvio padrão 2,3 Ω.

   Considere que as resistências têm distribuição normal.

   a. Determine um intervalo de predição de 95%; para a resistência de um único resistor.
   b. Determine um intervalo de tolerância para a resistência que inclua 90% dos resistores com confiança de 95%

2. (Seção 5.5- Exercício 2.) Em uma amostra de 20 parafusos, o torque de ruptura médio foi de 89,7 J com um desvio padrão de 8,2 J.

   a. Determine um intervalo de predição de 99%; para oo torque de ruptura de um único parafuso .
   b. Determine um intervalo de tolerância para o torque de ruptura que inclua 95% dos parafusos com confiança de 99%

3. (Seção 5.5- Exercício 3.)

   O artigo em Inglês "Ozone for Removal of Acute from LogyardRun-off"(Ozônio para Remoção de Toxidade Aguda de Água de Enxurrada de Pátios de Tora de Madeira)(M.Zenaits and S. Duff, Ozone Scienceand Engineering,2002:83-90) mostra as análises químicas deágua de enxurrada a partir de serrarias em British Columbia. Foram incluídas medições de PH para seis amostras de água:

$$5,9 \ ; 5,0 \ ; 6,5 \ ; \ 5,6 \ 5,9 \ ; \ 6,5.$$

Considerando que esses valores são um amostra aleatória de uma população normal.

a. Determine um intervalo de predição de 98%; para um pH de um único item da amostra.
b. Determine um intervalo de tolerância para o pH que inclua 95% dos itens da amostra com 95% de confiança.

4. (Seção 5.5- Exercício 4.)

   Seis medições foram feitas da concentração (em percentagem) de cinzas em determinada variedade de espinafre. A média amostral foi 19,35 e o desvio padrão foi de 0,577.

   Considere que as concentrações têm distribuição normal.

   a. Determine um intervalo de predição de 90%; para uma única medição. amostra.
   b. Determine um intervalo de tolerância para o teor de cinzas que inclui 99% das medições com 95% de confiança.

5. (Seção 5.4- Exercício 5.) Foram feitas cinco medições da taxa de octanagem para um tipo particular de gasolina. Os resultados em %foram.

$$77,0; 86,0; 86,5; 88,0; 85,3.$$

   a. Determine um intervalo de predição de 95%; para uma única medição. amostra.
   b. Determine um intervalo de tolerância para a taxa de octanagem que inclui 90% das medições com 99% de confiança.