Inferência II

Inferências com amostras pequenas

Felipe Figueiredo

Instituto Nacional de Traumatologia e Ortopedia

Sumário

Recapitulando

- Quando vamos fazer uma inferência sobre μ e sabemos σ^2, podemos usar σ diretamente no intervalo de confiança.
- Para isto, consultamos na tabela normal padrão (tabela Z) para obter o valor crítico z_c
- Esse valor crítico representa a probabilidade de que o intervalo criado em torno de $\hat{\mu} = \bar{x}$ contenha o valor desejado μ.
- Na prática, isso raramente acontece (se não sabemos μ, raramente saberemos σ^2).

Recapitulando

- Uma situação mais realista é quando queremos estimar μ e não sabemos σ.
- Quando temos uma amostra grande ($n \geq 30$), podemos aproximar σ por s, e usar s diretamente no cálculo da margem de erro
- Isso é justificado pelo Teorema Central do Limite (TCL) (e.g. vídeo do experimento de Galton).
- Consultamos o z_c na tabela Z, usando s como estimador de σ

A tabela Z

- Para a construção de intervalos de confiança, usamos o nível de confiança c (tipicamente $c = 0.95$).
- Isto é equivalente à significância $\alpha = 1 - 0.95 = 0.05$
- Isto é, a confiança ($c =$ probabilidade de que o IC contenha a média) é o complementar da significância ($\alpha =$ probabilidade de que o IC não contenha a média).
- Pela forma como a tabela é organizada, é mais conveniente procurar pela significância α na tabela.
- A significância deve ser dividida entre as duas caudas.

A tabela Z

Inferência II
Felipe Figueiredo
Recapitulando
Intervalos de confiança para a média
Resumo

- A tabela da Normal Padrão mostra os valores sob a curva até o ponto z observado (à esquerda de z).
- Cada linha corresponde ao primeiro dígito da área, e cada coluna identifica o segundo dígito da área (figura a seguir)

Example

A probabilidade de uma variável aleatória Z ser menor que z=0.35 é:

$$P(Z < 0.35) = 0.6368 = 63.68\%$$

A tabela Z

Inferência II
Felipe Figueiredo
Recapitulando
Intervalos de confiança para a média
Resumo

Tables of the Normal Distribution

Probability Content from -oo to Z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	[illegible]	0.6026	0.6064	0.6103	0.6141
0.3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	[illegible]	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

- $c = 95\% = 0.95$
- $\alpha = 5\% = 0.05$
- $\dfrac{\alpha}{2} = 2.5\% = 0.0250$
- $1 - 0.025 = 0.9750$
- Assim, o z_c é 1.96

E se a amostra não for grande?

Inferência II
Felipe Figueiredo
Recapitulando
Intervalos de confiança para a média
Resumo

- Quando a amostra é pequena, não podemos simplesmente substituir σ por s na fórmula, pois o erro dessa aproximação não é desprezível.
- Nesse caso, a média amostral não tem distribuição normal.
- Assim precisamos usar uma outra distribuição (tabelada) com a distribuição t de Student.

A distribuição t de Student

- Student (pseudônimo de W. S. Gossett [1876-1937], trabalhando para a cervejaria Guiness) criou uma distribuição que melhor se aproxima dos dados de amostras pequenas
- Tem um parâmetro graus de liberdade (gl) vinculado ao tamanho da amostra *n*.

A distribuição t de Student

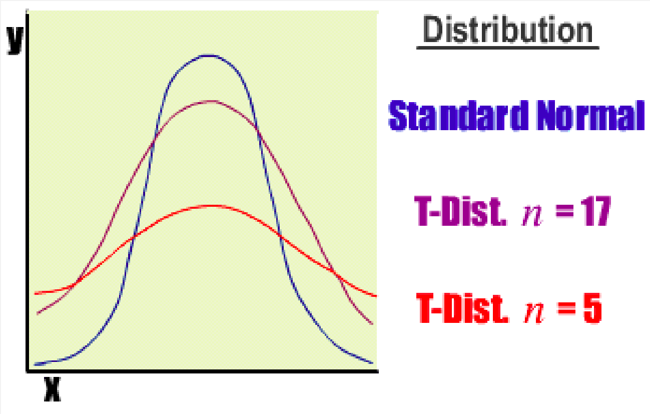

Figura: A distribuição t de Student

Propriedades da distribuição t

- A distribuição tem forma de sino (simétrica) assim como a Normal padrão Z
- Reflete a maior variabilidade inerente às amostras pequenas
- O formato da curva depende do tamanho da amostra n
- Quanto mais graus de liberdade (dados), mais a distribuição t se parece com a distribuição Z.

Intervalos de confiança para a média

Definition

A margem de erro usando a estatística t é

$$E = t_c \times \frac{s}{\sqrt{n}}$$

- Consultamos a tabela t de Student para encontrar o valor crítico t_c
- Graus de liberdade: $gl = n - 1$ (onde n é o tamanho da amostra)

A tabela t

t Distribution

	α					
Degrees of freedom	.005 (one tail) .01 (two tails)	.01 (one tail) .02 (two tails)	.025 (one tail) .05 (two tails)	.05 (one tail) .10 (two tails)	.10 (one tail) .20 (two tails)	.25 (one tail) .50 (two tails)
1	63.657	31.821	12.706	6.314	3.078	1.000
2	9.925	6.965	4.303	2.920	1.886	.816
3	5.841	4.541	3.182	2.353	1.638	765
4	4.604	3.747	2.776	2.132	1.533	741
5	4.032	3.365	2.571	2.015	1.476	727
6	3.707	3.143	2.447	1.943	1.440	.718
7	3.500	2.998	2.365	1.895	1.415	.711
8	3.355	2.896	2.306	1.860	1.397	.706
9	3.250	2.821	2.262	1.833	1.383	.703
10	3.169	2.764	2.228	1.812	1.372	.700
11	3.106	2.718	2.201	1.796	1.363	.697
12	3.054	2.681	2.179	1.782	1.356	.696
13	3.012	2.650	2.160	1.771	1.350	.694
14	2.977	2.625	2.145	1.761	1.345	.692
15	2.947	2.602	2.132	1.753	1.341	.691
16	2.921	2.584	2.120	1.746	1.337	.690
17	2.898	2.567	2.110	1.740	1.333	.689
18	2.878	2.552	2.101	1.734	1.330	.688
19	2.861	2.540	2.093	1.729	1.328	.688
20	2.845	2.528	2.086	1.725	1.325	.687
21	2.831	2.518	2.080	1.721	1.323	.686
22	2.819	2.508	2.074	1.717	1.321	.686
23	2.807	2.500	2.069	1.714	1.320	.685
24	2.797	2.492	2.064	1.711	1.318	.685
25	2.787	2.485	2.060	1.708	1.316	.684
26	2.779	2.479	2.056	1.706	1.315	.684
27	2.771	2.473	2.052	1.703	1.314	.684
28	2.763	2.467	2.048	1.701	1.313	.683
29	2.756	2.462	2.045	1.699	1.311	.683
Large (z)	2.575	2.327	1.960	1.645	1.282	.675

Exemplo

Example

Considere uma amostra de 10 bebês selecionada de uma população de bebês que recebe antiácidos que contém alumínio e são frequentemente usados para tratar distúrbios digestivos. A distribuição de níveis de alumínio no plasma é conhecida como aproximadamente normal, no entanto sua média e desvio padrão não são conhecidos. O nível médio de alumínio para a amostra de dez bebês é 37.2 μg/l e desvio-padrão 7.13 μg/l. Calcule um intervalo com 95% de confiança para a média populacional.

(Fonte: Hacker & Simões, 2008, Fiocruz)

Exemplo

- $\bar{x} = 37.2$
- $s = 7.13$
- $n = 10 \Rightarrow gl = 9$

Solução

$$t_c = 2.262$$

$$E = t_c \times \frac{s}{\sqrt{n}}$$

$$E = 2.262 \times \frac{7.13}{\sqrt{10}} \approx 5.1$$

$$IC(95\%) = (37.2 - 5.1, 37.2 + 5.1) = (32.1, 42.3)$$

Exercício

Exercício

Num estudo para descrever o perfil dos pacientes adultos atendidos no ambulatório de um posto de saúde, uma amostra de 16 pacientes adultos foi selecionada ao acaso entre o total de pacientes atendidos no posto durante os últimos três anos, coletando-se dos prontuários desses pacientes dados relativos à idade, à escolaridade e a outros fatores de interesse.

Para a variável idade, observou-se uma média amostral de 36.86 anos com um desvio padrão amostral de 17.79 anos.

Exercício

Exercício

1. Defina a população e a amostra.
2. Forneça uma estimativa pontual, um intervalo de 90% de confiança e um intervalo de 95% de confiança para a idade média dos adultos atendidos neste ambulatório nos últimos três anos. Interprete e compare os intervalos de confiança.

$$E = \frac{t_c s}{\sqrt{n}} \qquad \bar{x} = 36.86$$

$$t_c(90\%) = 1.753 \qquad s = 17.79$$

$$t_c(95\%) = 2.132 \qquad n = 16 \Rightarrow gl = 15$$

Exercício

Inferência II
Felipe Figueiredo
Recapitulando
Intervalos de confiança para a média
A distribuição t de Student
Intervalos de confiança para amostras pequenas
Resumo

Solução

- IC de 90% (c=0.90)

$$E = \frac{t_c s}{\sqrt{n}} = \frac{1.753 \times 17.79}{\sqrt{16}} \approx 7.80$$

$$IC_{0.90} = \bar{x} \pm E = 36.86 \pm 7.80 = (29.06, 46.66)$$

- IC de 95% (c=0.95)

$$E = \frac{t_c s}{\sqrt{n}} = \frac{2.132 \times 17.79}{\sqrt{16}} \approx 9.48$$

$$IC_{0.95} = \bar{x} \pm E = 36.86 \pm 9.48 = (27.38, 46.34)$$

Resumo

Inferência II
Felipe Figueiredo
Recapitulando
Intervalos de confiança para a média
Resumo

Para construir um intervalo de confiança para a média μ devemos considerar as informações e dados disponíveis:

- Se soubermos σ, usamos a tabela Z (z_c)

$$E = z_c \frac{\sigma}{\sqrt{n}}$$

- Se não soubermos σ, mas se n é grande ($n \geq 30$), usamos a tabela Z (z_c)

$$E = z_c \frac{s}{\sqrt{n}}$$

- Se não soubermos σ, mas e se n é pequeno ($n < 30$), usamos a tabela t (t_c)

$$E = t_c \frac{s}{\sqrt{n}}$$