ESTIMAÇÃO DA MÉDIA DE UMA DISTRIBUIÇÃO NORMAL (OU APROXIMADAMENTE NORMAL)

Na estimação da proporção era

$$\gamma = \mathbb{P}\left[-z \leq Z \leq z\right], \qquad Z \sim \mathcal{N}(0, 1^2)$$

Agora

$$\gamma = \mathbb{P}\left[-z \leq T_{n-1} \leq z\right],$$

onde T_{n-1} é uma variável aleatória cujos valores são tabeladas (veja a tabela abaixo) chamada T de Student com $n-1$ graus de liberdade; n usado aquí é o tamanho de amostra.
O pressuposto inicial de que toda a população possui a distribuição aproximadamente normal é o ingrediente necessário para o uso da distribuição de Student nos moldes a serem apresentados abaixo.

Amostra de $n = 13$ observações do tempo de rodagem de pneus:

$$1, 1.4, 2, 2.4, 2.7, 2.9, 3.1, 3.5, 3.9, 4, 4.6, 5.2, 6.1$$

A média dessa amostra

$$\bar{x}_{13} \equiv \hat{\mu} = \frac{1 + 1.4 + \cdots + 6.1}{13} = 3,292308 \approx 3,3$$

toma-se como a estimativa pontual de μ, a média populacional.

A variância dessa amostra

$$s_x^2 = \frac{1}{n-1} \left(\sum_{i=1}^{13} x_i^2 \; - \; n(\bar{x})^2 \right) = 2,165769 \approx 2,17$$

e o desvio padrão $s_x = \sqrt{2,165769} \approx 1,47$ tomam-se como as estimativas pontuais para a variância (σ^2) e o desvio padrão (σ) populacionais.

Se deseja-se fazer a estimativa intervalar para μ com o coeficiente de confiança $\gamma = 0,8$ então, em primeiro lugar tem que achar z usando a tabela abaixo.

Toma-se a linha 12 pois no nosso caso, o número de graus de liberdade é $13 - 1$ ("graus de liberdade" é igual, por definição, ao "tamanho de amostra" -1). Depois, toma-se a coluna "Two tails 0,20" por que nas duas caudas tem que ter 0,20 de probabilidade um vez que queremos que na parte central esteja 0,8 da probabilidade. Alias, a última linha da tabela apresenta diretamente aquilo que precisamos: o desejado valor do coeficiente de confiança. Na intersecção entre a linha e a coluna escolhidas, há o valor que nos interessa: $z = 1,356$.

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%

O intervalo de confiança tem o seguinte formato

$$\left[\hat{\mu} - z\sqrt{\frac{s_x^2}{n}};\ \hat{\mu} + z\sqrt{\frac{s_x^2}{n}};\right]$$

ou

$$\left[\hat{\mu} - z\frac{s_x}{\sqrt{n}};\ \hat{\mu} + z\frac{s_x}{\sqrt{n}};\right]$$

No nosso caso:

$$[3,3 - 0,5528475;\ \ 3,3 + 0,5528475]$$

Exercícios sobre o tema “estimação de média”

Exc. 1. Podemos assumir que o tempo de rodagem de pneus de uma certa marca tenha distribuição aproximadamente normal. Estime, por intervalo de confiança com o coeficiente de confiança $0,9$, a média dessa distribuição usando a seguinte amostra que apresenta o tempo de rodagem (em anos) de 10 pneus da marca escolhidos aleatoreamente:

$$2.21, 2.03, 1.75, 1.50, 1.65, 1.72, 1.33, 2.96, 1.53, 1.40$$

Exc. 2. 101 medições do peso de peixe de uma certa espécie dão os seguintes resultados:

$$\sum_{i=1}^{101} x_i = 235, \qquad \sum_{i=1}^{101} x_i^2 = 573$$

Assuma que o peso de peixe desta espécie tenha distribuição aproximadamente normal. Use os dados fornecidos para estimar a média dessa distribuição por intervalo de confiança com o coeficiente de confiança $0,98$. Faça a outra estimativa, agora com o coeficiente de confiança $0,7$.

Ex.1

Que seja claro: “1kg” é o que está escrito nos pacotes de cafe recebidos por supermercado, mas o supermecado não tem certeza que isto é verdade, quer dizer, que a máquina que encha pacotes foi de fato regulada para 1kg.
Por não saber a regulagem da máquina, o peso média de pacotes de café, μ, é um incognito. Porém, devido a processo de enchimento, sabemos que para qualquer valor de μ, o desvio padrão do peso de pacotes de café é $\sigma = 0,01$Kg. Em outras palavras, sabemos o desvio padrão populacional mas não sabemos a média populacional e queremos estimar esta por intervalo de confiança.

No item **(a)** do exercício, o coeficiente de confiança da estimativa deve ser de 90%, e a precisão da estimativa deve ser $0,005$Kg. Recorde da página 16 da aula que a estimativa intervalar terá a "cara":

$$\left[\bar{X}_n - \varepsilon \,;\, \bar{X}_n + \varepsilon\right]$$

onde $\bar{X}_n$ denota a média amostral a ser construida com base na amosta de tamanho n, e

$$\varepsilon = \frac{z\sigma}{\sqrt{n}}, \text{ sendo que } \mathbb{P}\left[-z \leq Z \leq z\right] = \gamma.$$

No nosso caso, $\gamma = 0,9$, de onde (usando a tabela da distribuição normal padrão) $z = 1,65$, e $\varepsilon = 0,005$. Portanto, o desconhecido n deve satisfazer a relação

$$0,005 = \frac{1,65 \cdot 0,01}{\sqrt{n}} \;\Leftrightarrow n = \frac{0,0165^2}{0,005^2} = (3,3)^2 \approx 11.$$

(b) A diferença do item (a) está somente no valor de γ, que era de 90% e ficou agora em 95%. Isto afeta o valor de z: no item (a) era $z_{\gamma=0,9} = 1,65$, enquanto que agora $z_{\gamma=0,95} = 1,96$. Repetindo os argumentos do item (a), temos que

$$n = \left(\frac{z_\gamma \sigma}{\varepsilon}\right)^2 = \left(\frac{1,96 \cdot 0,01}{0,005}\right)^2 \approx 16.$$

(c) Aquí, a situação é pouco diferente daquela que era em comum nos itens (a) e (b). Agora a amostragem foi feita: $n = 100$ pacotes foram pesados. Deseja-se então saber a confiança (γ) do intervalo de confiança para a média populacional cuja precisão é de 0,003 (isto é: $\varepsilon = 0,003$). Observe que o resultado da amostragem, apropreadamente dito, não foi revelado pelo enunciado, isto é, não foi dado para nos o valor que $\bar{X}_{n=100}$ assumiu. Mas tambem, não precisamos deste valor para responder na pergunta colocada, pois da fórmula $\varepsilon = \frac{z\sigma}{\sqrt{n}}$ deduz-se que

$$z = \frac{\varepsilon\sqrt{n}}{\sigma} = \frac{0,003\sqrt{100}}{0,01} = 3.$$

De onde, com o uso da tabela da distribuição Normal Padrão, segue-se que

$$\gamma = \mathbb{P}\left[-3 \leq Z \leq 3\right] = 0,9973.$$

Observação: Já comentamos que o tamanho do intervalo de confiança $\left[\bar{X}_n - \varepsilon\,;\bar{X}_n + \varepsilon\right]$ é 2ε, mas que a precisão da estimativa é ε e isto ocorre pois por "precisão" entende-se a diferença entre o valor estimado μ por sua estimativa $\bar{X}_n$. Tendo em mente esta diferença, é natural perguntar se o valor de ε desejado seria $0,003/2$ em vez de $0,003$. É a questão de interpretação. Na minha interpretação, o quem compus o exercício equivocou-se e escreveu "tamanho" onde quiz "precisão".

Ex.3.

(a) Pelo enunciado, $\sigma = 0,1$ m, quer dizer, o desvio padrão populacional da altura de toda a população é 0,1 m, valendo para quelquer que seja a altura média populacional.

Neste item, sabemos que $n = 30$ e que $\bar{X}_{n=30}$ assumiu valor 1,45 m, e queremos contruir o intervalo de confiança com $\gamma = 0,95$ (coeficiente de confiança igual a 95%, quer dizer). Já fizemos as contas que mostram que $z_{\gamma=0,95} = 1,96$. Portanto,

$$\varepsilon = \frac{z_{\gamma=0,95}\sigma}{\sqrt{n}} = \frac{1,96 \cdot 0,1}{\sqrt{30}} = 0,035.$$

Portanto, o intervalo de confiança para a média tem a seguinte forma:

$$[1,45 - 0,035\,;\; 1,45 + 0,035]\,.$$

(b) Neste item queremos estimar a média populacional por intervalo de confiança com $\varepsilon = 0,05$ m e $\gamma = 99\%$. A pergunta é qual deve ser o temanho de amostra para que a precisão e a confiança desejadas sejam garantidas. Já tivemos um problema parecido e sabemos que para sua solução é precisa achar z_γ e usar a fórmula $n = \left(\frac{z_\gamma \sigma}{\varepsilon}\right)^2$. Então:

$$n = \left(\frac{z_{\gamma=0,99}\sigma}{\varepsilon}\right)^2 = \left(\frac{2,58 \cdot 0,1}{0,05}\right)^2 = 26,6256 \approx 27.$$

Ex. 2.

(a) Do enunciado, temos que $n = 100$, que o valor média da amostra (valor de $\bar{X}_{n=100}$) é 20, que a desvio padrão da amostra é 6 (isto é, $s = 6$), e que $\varepsilon = 1$ (pois 2ε corresponde ao tamanho do intervalo de confiança, e sendo que o tamanho foi de $21 - 19 = 2$ minutos, concluimos que $\varepsilon = 1$). A fórmula que vincula todos estes valores (menos o de $\bar{X}_{n=100}$) com z_γ é:

$$z_\gamma = \frac{\varepsilon\sqrt{n}}{s}$$

Logo,

$$z_\gamma = \frac{1 \cdot \sqrt{100}}{6} = 1,66$$

o que dá, via a tabela da distribuição Normal Padrão, que

$$\gamma = 0,90308.$$

(b) Já que este item pergunta sobre o tamanho de amostra, entendemos que tal amostra não foi feita, e portanto, não tem como saber nem a média amostral nem o desvio padrão amostral (s). O desvio padrão populacional (σ) tambem não foi dado. Sem saber s ou σ, não há como usar nossas fórmulas. Portanto, a única maneira de achar alguma solução do problema, é aproveitar do resultado da amostragem feita com 100 pacientes para usar s como estimativa para o desvio pasdrão amostral. Então, assumiremos que $\sigma = 6$. Temos do enunciado que $\varepsilon = 1$ e que $\gamma = 95\%$. Já sabemos que $z_{\gamma=0,95} = 1,65$. Portanto,

$$n = \left(\frac{z_{\gamma=0,95} \cdot \sigma}{\varepsilon}\right)^2 = \left(\frac{1,65 \cdot 6}{1}\right)^2 = 98,01 \to 99.$$