Métodos Quantitativos

Prof. Júlio Cesar Nievola

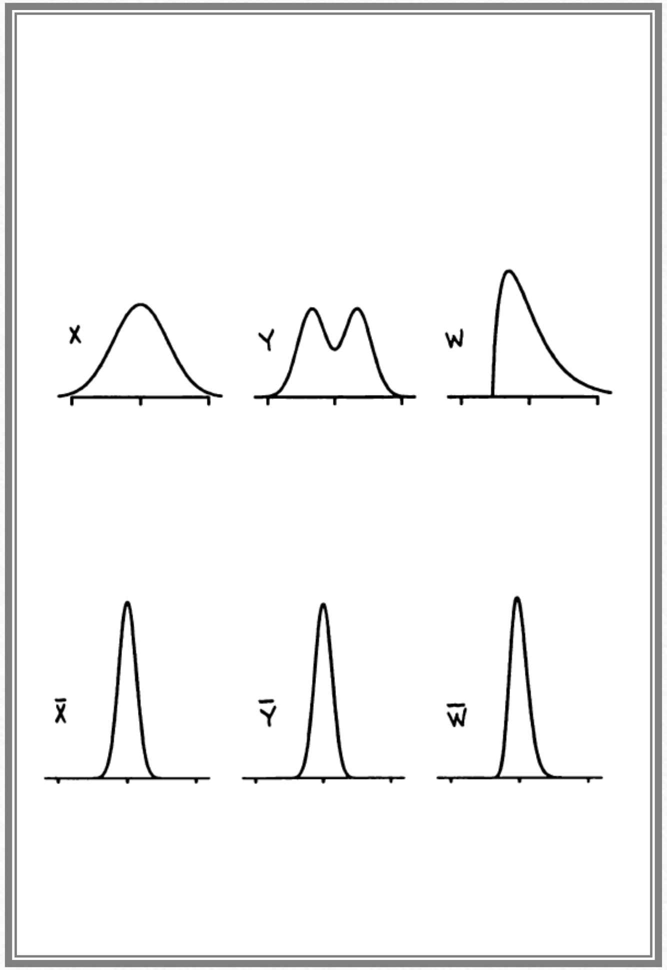
PPGIa – PUCPR

Teorema do Limite Central

- Se forem obtidas amostras aleatórias de tamanho n de uma população com média μ e desvio padrão σ, à medida que n aumenta $\overline{X}$ se aproxima da distribuição normal com média μ e desvio padrão ${}^{\sigma}/_{\sqrt{n}}$. Então:

$$Pr\left(a \leq \overline{X} \leq b\right) = Pr\left(\frac{a-\mu}{{}^{\sigma}/_{\sqrt{n}}} \leq Z \leq \frac{b-\mu}{{}^{\sigma}/_{\sqrt{n}}}\right)$$

Importância do Teorema do Limite Central



- Independente do formato da distribuição original, o uso de médias resulta em uma normal. Para encontrar a distribuição de $\overline{X}$ é necessário somente a média e o desvio padrão da população.

- As três densidades de probabilidade ao lado (X, Y e W) tem a mesma média e desvio padrão. Apesar da diferença nos formatos, quando $n = 10$ as distribuições amostrais da média, $\overline{X}$, $\overline{Y}$ e $\overline{W}$ são aproximadamente idênticas.

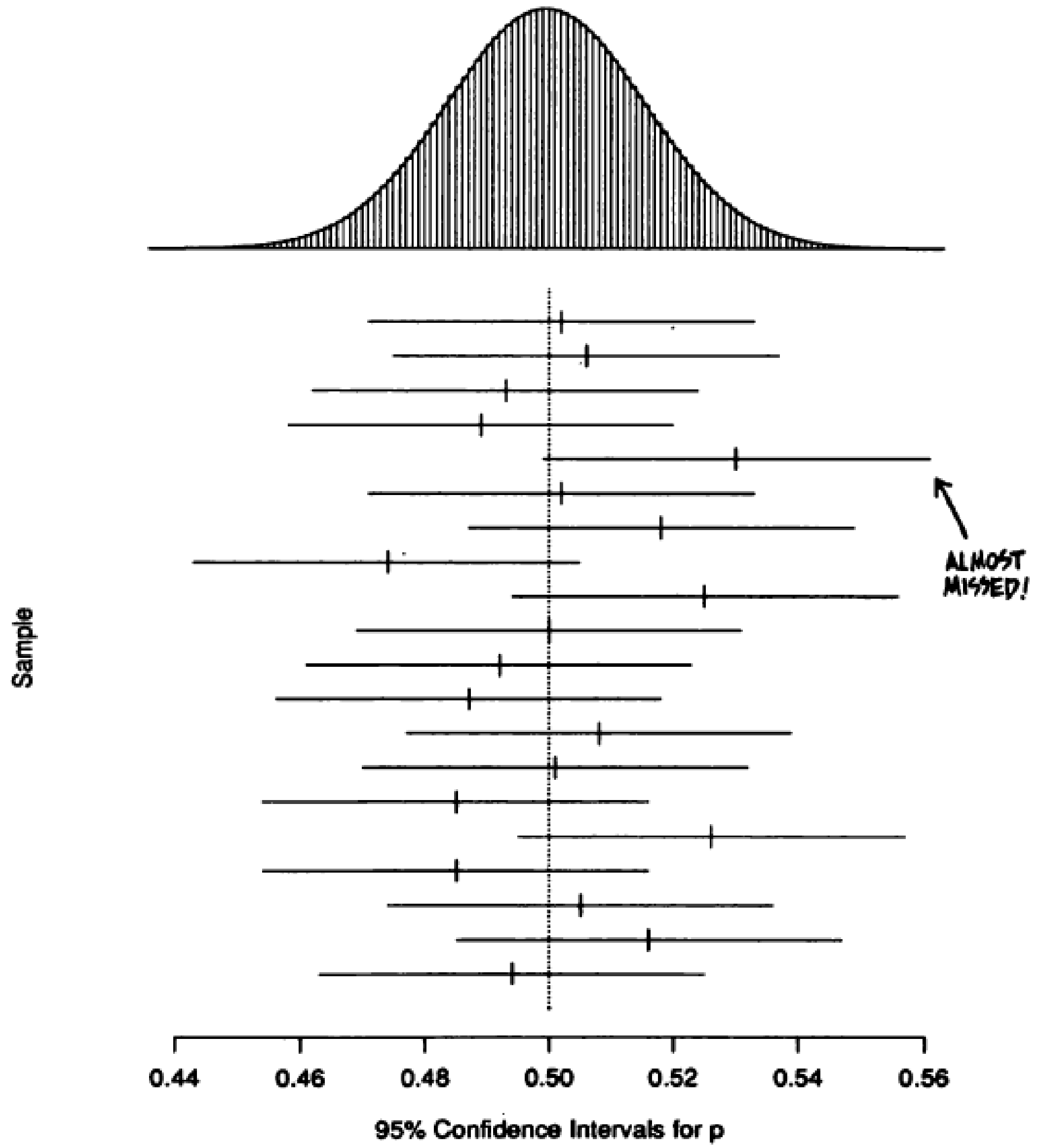
Intervalo e Nível de Confiança

- Um intervalo de confiança é uma faixa de valores usado para estimar um parâmetro de uma população (por exemplo uma proporção ou uma média) e está associado a um nível de confiança específico.

- Um nível de confiança é a probabilidade da estimativa do intervalo incluir o parâmetro da população (por exemplo a proporção ou a média).

Resultados de uma Simulação de 20 amostras com valor real p = 0,5.

Abaixo intervalos de confiança de 95%



Relação entre Intervalo e Nível de Confiança

- O aumento do nível de confiança torna o intervalo de confiança maior e menos preciso.

- O aumento do tamanho da amostra reduz a largura do intervalo de confiança, o que aumenta a precisão.

Exemplo: Intenções de Voto em Eleição – 1

- O candidato a senador XYMK?T deseja saber qual proporção do eleitorado irá votar nele nas próximas eleições e para saber isto ele contratou uma empresa de pesquisas. A empresa entrevistou n = 1.000 eleitores, descobrindo que 550 eleitores tinham intenção de votar no candidato XYMK?T, ou seja, $\hat{p} = 0{,}55$. Quando a empresa apresentou estes resultados o candidato respondeu: "Vocês entrevistaram apenas 1.000 eleitores; mas existem muitos milhões de eleitores, somente no meu estado! Qual a confiança que posso ter neste resultado?". A empresa respondeu: "Podemos afirmar com 95% de confiança que a quantidade de votos estará entre 51,9% e 58,1%".

Exemplo: Intenções de Voto em Eleição – 2

- A distribuição de $\hat{p}$ é aproximadamente normal com média p e desvio padrão dado por:

$$\sigma(\hat{p}) = \frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

- Como a curva é normal usa-se a transformada Z e uma tabela padrão para encontrar a largura do intervalo dentro do qual 95% dos valores estarão. Esta largura corresponde a 1,96 desvios-padrão, ou seja, tem-se:

$$0{,}95 = Pr(-1{,}96 \leq Z \leq 1{,}96)$$

Exemplo: Intenções de Voto em Eleição – 3

- De acordo com a definição da transformada Z

$$0{,}95 \cong Pr\left(-1{,}96 \leq \frac{\hat{p} - p}{\sigma(p)} \leq 1{,}96\right)$$

- Isto se torna

$$0{,}95 \cong Pr\big(p - 1{,}96\sigma(p) \leq \hat{p} \leq p + 1{,}96\sigma(p)\big)$$

- Ou seja

$$0{,}95 \cong Pr\big(\hat{p} - 1{,}96\sigma(p) \leq p \leq \hat{p} + 1{,}96\sigma(p)\big)$$

Exemplo: Intenções de Voto em Eleição – 4

- Como não se tem o valor de $\sigma(p)$ usa-se a aproximação dada pelo erro padrão (SE) de $\hat{p}$ que é

$$SE(\hat{p}) = \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

- Com isto tem-se a fórmula

$$0{,}95 \cong Pr\big(\hat{p} - 1{,}96 \cdot SE(\hat{p}) \leq p \leq \hat{p} + 1{,}96 \cdot SE(\hat{p})\big)$$

- Portanto, com 95% de confiança, a quantidade de votantes no senador XYMK?T estará na faixa

$$\hat{p} \pm 1{,}96 \cdot SE(\hat{p}) = 0{,}55 \pm 1{,}96 \cdot \sqrt{\frac{0{,}55 \cdot (1-0{,}55)}{1.000}}, ou\ seja, (55 \pm 3{,}1)\%.$$

Estimativa da Proporção da Amostra

- Usa-se a distribuição normal para construir um intervalo de confiança em torno da proporção da amostra quando $np \geq 5$ e $nq \geq 5$.

Estatística da População	Distribuição da População	Condições	Intervalo de Confiança
p	Binomial	n ser grande p_s *ser a proporção na amostra* q_s *ser* $1 - p_s$	$\left(p_s - c \cdot \sqrt{\frac{p_s q_s}{n}}, p_s + c \cdot \sqrt{\frac{p_s q_s}{n}}\right)$

Nível de Confiança	Valor de c
90%	1,64
95%	1,96
99%	2,58

Cálculo do Intervalo de Confiança com σ^2 conhecido

Estatística da População	Distribuição da População	Condições	Intervalo de Confiança
μ	Normal	Conhecer σ^2 n pode ser grande ou pequeno $\overline{x}$ é a média da amostra	$\left(\overline{x} - c \cdot \frac{\sigma}{\sqrt{n}}, \overline{x} + c \cdot \frac{\sigma}{\sqrt{n}}\right)$
μ	Não normal	Conhecer σ^2 n é grande (pelo menos 30) $\overline{x}$ é a média da amostra	$\left(\overline{x} - c \cdot \frac{\sigma}{\sqrt{n}}, \overline{x} + c \cdot \frac{\sigma}{\sqrt{n}}\right)$
μ	Normal ou não normal	Não se conhece σ^2 n é grande (pelo menos 30) $\overline{x}$ é a média da amostra s^2 é a variância da amostra	$\left(\overline{x} - c \cdot \frac{s}{\sqrt{n}}, \overline{x} + c \cdot \frac{s}{\sqrt{n}}\right)$

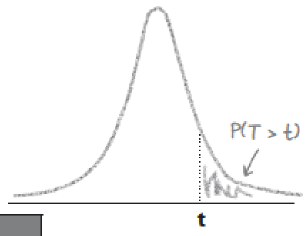
Pequenas amostras com σ desconhecido

- Usa-se a distribuição t para construir um intervalo de confiança quando a população segue a distribuição normal (ou aproximadamente normal), o tamanho da amostra é menor que 30 e o desvio padrão da população, σ, é desconhecido.

Estatística da População	Distribuição da População	Condições	Intervalo de Confiança
μ	Normal ou não normal	σ^2 é desconhecido n é pequeno (menor que 30) $\overline{x}$ é a média da amostra s^2 é a variância na amostra	$\left(\overline{x} - t(\vartheta) \cdot \frac{s}{\sqrt{n}}, \overline{x} + t(\vartheta) \cdot \frac{s}{\sqrt{n}}\right)$

*2. t-distribution critical values

This table gives you the values of t where P(T > t) = p. T follows a t-distribution with v degrees of freedom. Look up the values of v and p and look up t.



Look up v in the first column... ...look up p in the first row...

t

	Tail probability p											
v	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

...then read off the value of t from the table.

Exemplo – Pequenas amostras

- Qual é o intervalo de confiança, com 95% de nível de confiança, para o peso de balas de goma quando se obteve 10 exemplos com $\overline{x} = 0{,}5$g e $s^2 = 0{,}09$?
- Como são 10 exemplos na amostra o número de graus de liberdade é 9. Para encontrar o intervalo com 95% de confiança busca-se 0,025 com nove graus de liberdade na tabela e encontramos *t=2,262*. Substituindo na fórmula:

$$(v_{mín}, v_{máx}) = \left(0.5 - 2.262 \cdot \sqrt{{}^{0.09}/_{10}}, 0.5 + 2.262 \cdot \sqrt{{}^{0.09}/_{10}}\right)$$

$$(v_{mín}, v_{máx}) = (0.285\,, 0.715)$$