

ACH3584 - Estatística II

Aula 04: Inferência Estatística

Alexandre Ribeiro Leichsenring

alexandre.leichsenring@usp.br

Organização

- 1 Estatísticas e Parâmetros
- 2 Distribuições Amostrais
- 3 Distribuição Amostral da Média

Organização

- 1 Estatísticas e Parâmetros
- 2 Distribuições Amostrais
- 3 Distribuição Amostrai da Média

Amostra

Definição: Uma amostra aleatória simples de tamanho n de uma variável aleatória X é o conjunto de n variáveis independentes $X_1, X_2, \dots, X_n$, cada uma com a mesma distribuição de X .

Ou seja, a amostra será a n -upla ordenada $(X_1, X_2, \dots, X_n)$, em que X_i indica a observação do i -ésimo elemento sorteado.

Exemplo

Vamos retirar uma AAS de 5 alturas (em cm) de uma população de mulheres cujas alturas X seguem a distribuição Normal com média 167 e desvio padrão 5, isto é,

$$X \sim N(167; 5^2)$$

Um possível resultado da extração dessa amostra é dada por:

$$x_1 = 165 \quad x_2 = 161 \quad x_3 = 168 \quad x_4 = 173 \quad x_5 = 173$$

Estatísticas e parâmetros

Obtida uma amostra, desejamos usá-la para produzir informação sobre alguma característica específica.

Por exemplo, podemos querer calcular a média da amostra $(X_1, X_2, \dots, X_n)$:

$$\bar{X} = \frac{1}{n} \{X_1 + X_2 + \dots + X_n\}$$

► $\bar{X}$ também é uma variável aleatória.

Qualquer outra característica da amostra será sempre uma função do vetor aleatório $(X_1, X_2, \dots, X_n)$.

Definição. Uma *estatística* é uma característica da amostra. Ou seja, é uma função de $X_1, X_2, \dots, X_n$.

Estatísticas mais comuns

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad : \quad \text{média da amostra;}$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad : \quad \text{variância da amostra;}$$

$$X_{(1)} = \min(X_1, X_2, \dots, X_n) \quad : \quad \text{o menor valor da amostra;}$$

$$X_{(n)} = \max(X_1, X_2, \dots, X_n) \quad : \quad \text{o maior valor da amostra;}$$

$$W = X_{(n)} - X_{(1)} \quad : \quad \text{amplitude da amostra.}$$

Características da população x Características da amostra

Definição. Um *parâmetro* é uma medida usada para descrever uma característica da população.

Denominação	População	Amostra
Núm. elementos	N	n
Média	$\mu = E(X)$	$\bar{X} = \frac{\sum X_i}{n}$
Variância	$\sigma^2 = \text{Var}(X)$	$S^2 = \frac{\sum (X_i - \bar{X})^2}{(n-1)}$
Proporção	p	$\hat{p}$
Função densidade	$f(x)$	histograma

Organização

- 1 Estatísticas e Parâmetros
- 2 Distribuições Amostrais**
- 3 Distribuição Amostrai da Média

Distribuições Amostrais

- O problema da Inferência é fazer afirmações sobre a distribuição da população através da amostra
- Em particular, fazer afirmações sobre parâmetros que descrevem a população;
- Suponha que estamos interessados em uma determinada característica da população, digamos a média (μ):

Nossa decisão será baseada no valor de

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Colhida uma amostra e observados os valores

$$(x_1, x_2, \dots, x_n)$$

faremos nossa afirmação sobre a média populacional μ .



De maneira mais geral:

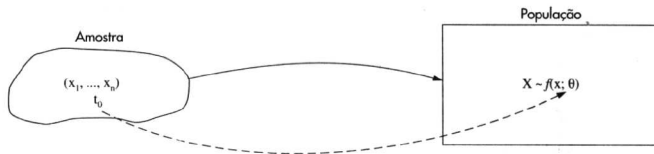
- Temos uma amostra

$$(X_1, X_2, \dots, X_n)$$

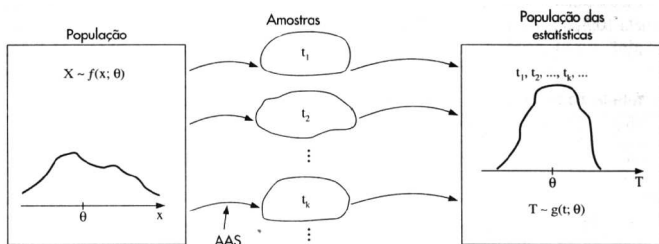
e uma estatística

$$T = f(X_1, X_2, \dots, X_n)$$

- Colhida a amostra, teremos observado um valor particular de T , digamos t_0
- Baseado nesse valor faremos a afirmação sobre o parâmetro populacional θ



- Validade da nossa resposta será melhor compreendida se soubermos o que ocorre com a estatística T , quando observamos todas as amostras “possíveis” de uma população
- Em outras palavras: qual a distribuição de T quando $(X_1, X_2, \dots, X_n)$ assume todos os valores possíveis?
- Essa distribuição é chamada de distribuição amostral da estatística T



Organização

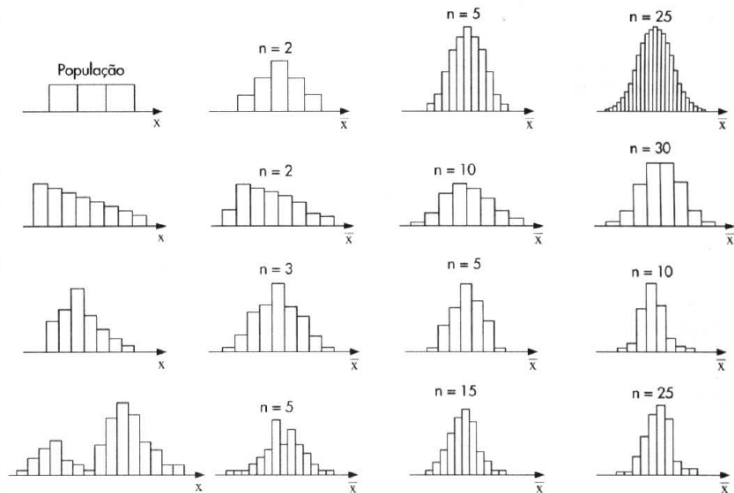
- 1 Estatísticas e Parâmetros
- 2 Distribuições Amostras
- 3 Distribuição Amostral da Média**

Distribuição Amostral da Média

Se X é uma v.a. com média μ e variância σ^2 , e seja $(X_1, X_2, \dots, X_n)$ uma amostra aleatória simples de X , então:

$$\begin{aligned}\mathbf{E}(\bar{X}) &= \mu \\ \text{Var}(\bar{X}) &= \frac{\sigma^2}{n}\end{aligned}$$

Distribuição amostral da média



Teorema do Limite Central

Para amostras aleatórias simples $(X_1, X_2, \dots, X_n)$ de uma população com média μ e variância σ^2 , a distribuição amostral da média $\bar{X}$ se aproxima, para n grande, de uma distribuição Normal com média μ e variância $\frac{\sigma^2}{n}$.

Exemplo

Voltemos ao exemplo das máquinas de encher pacotes de café cujos pesos (X) seguem uma distribuição $N(500, 100)$. Colhendo uma amostra de tamanho $n = 100$ pacotes e pesando-os, qual será a distribuição de $\bar{X}$?

Solução

Pelo que foi dito acima, $\bar{X}$ terá distribuição normal com:

$$\begin{aligned} \mathbf{E}(\bar{X}) &= \mu = 500 \\ \text{Var}(\bar{X}) &= \frac{\sigma^2}{n} = \frac{100}{100} = 1 \end{aligned}$$

Ou seja,

$$\bar{X} \sim N(500, 1)$$

Exemplo

Com relação ao exemplo anterior, retirando-se uma amostra de tamanho 100, responda:

- a) Qual a probabilidade de que a média $\bar{X}$ seja maior do que 502?
 $P(\bar{X} > 502) = ?$
- b) Qual a probabilidade de que a média $\bar{X}$ seja menor do que 498?
 $P(\bar{X} < 498) = ?$
- c) Qual a probabilidade de que a média $\bar{X}$ esteja entre 498 e 502?
 $P(498 < \bar{X} < 502) = ?$

Corolário

Se $X_1, X_2, \dots, X_n$ for uma amostra aleatória simples da população X com média μ e variância σ^2 e $\bar{X} = \frac{X_1 + \dots + X_n}{n}$, então:

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

Exemplo

Uma v.a. X tem distribuição normal, com média 100 e desvio padrão 10.

- a) Qual a $\mathbf{P}(90 < X < 110)$?
- b) Se $\bar{X}$ for a média de uma amostra de 16 elementos retirados dessa população, calcule $\mathbf{P}(90 < \bar{X} < 110)$.
- c) Represente num único gráfico a distribuição de X e de $\bar{X}$.

Exercício

A máquina de empacotar um determinado produto o faz segundo uma distribuição normal com média 513 e desvio padrão 10 g.

- a) Qual proporção de pacotes terá menos do que 500 g?
- b) Qual a probabilidade de que a média dos pesos de 4 pacotes selecionados ao acaso seja inferior a 500 g?

B.1 Normal Padrão

Valores de p tais que $P(Z < z_c) = p$.



Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,10	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,20	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,30	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,40	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,50	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,60	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,70	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,80	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,90	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,00	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,10	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,20	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,30	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,40	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,50	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,60	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,70	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,80	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,90	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,00	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,10	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,20	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,30	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,40	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,50	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,60	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,70	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,80	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,90	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,00	0,9987	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990
3,10	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,20	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,30	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,40	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998