

Técnicas de Amostragem - Aula 03

Amostragem Aleatória Simples Com Reposição (AASc)

Kaique Matias de Andrade Roberto

Ciências Atuariais

HECSA - Escola de Negócios

FIAM-FAAM-FMU

1. Conceitos que aprendemos em Aulas anteriores
2. Cálculos com a Distribuição Normal
3. Amostragem Aleatória Simples
4. Amostragem Aleatória Simples Com Reposição (AASc)
5. Normalidade e Intervalo de Confiança
6. Determinação do Tamanho da Amostra
7. Estimação da Proporção
8. Comentários Finais
9. Referências

Conceitos que aprendemos em Aulas anteriores

Conceitos que aprendemos em Aulas anteriores

Nas Aulas anteriores nós lidamos com as principais definições, nomenclaturas e terminologias da Teoria de Amostragem. A saber:

- população;
- amostra;
- planejamento amostral;
- estatísticas;
- distribuições amostrais;
- estimadores e suas propriedades.

Cálculos com a Distribuição Normal

A distribuição normal, também conhecida como distribuição Gaussiana, é uma das distribuições de probabilidade mais utilizadas, por pelo menos 2 motivos:

- permite modelar fenômenos naturais, estudos do comportamento humano, processos industriais, entre outros;

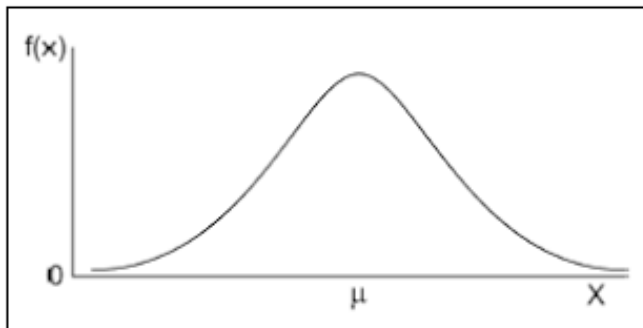
- possibilita o uso de aproximações para o cálculo de probabilidades de muitas variáveis aleatórias.

Definição 2.1

Uma variável aleatória X com média $\mu \in \mathbb{R}$ e desvio padrão $\sigma > 0$ tem **distribuição Normal** denotada $X \sim N(\mu, \sigma^2)$, se a sua função de distribuição de probabilidades for dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}.$$

Cálculos com a Distribuição Normal



Do gráfico da densidade da Normal podemos concluir algumas propriedades básicas:

- $f(x)$ é simétrica em relação a μ ;
- $f(x)$ decresce a medida que $|x|$ cresce;
- o valor máximo de $f(x)$ se dá para $x = \mu$.

Note que, para calcular $P(a \leq X \leq b)$ devemos realizar a conta

$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

Também pode-se demonstrar que se $X \sim N(\mu, \sigma^2)$ então

$$E(X) = \mu \text{ e } \text{Var}(X) = \sigma^2.$$

Claro, vocês em geral não fizeram cursos de Cálculo Diferencial e Integral. Mas mesmo se esse fosse o caso, essa integral **não é calculável (não existe fórmula fechada)**.

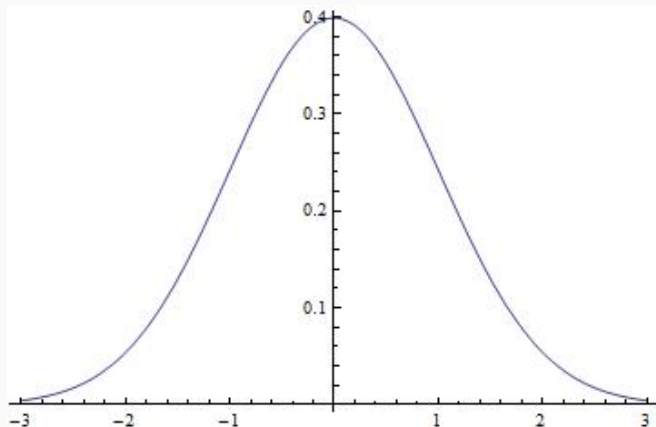
Apesar disso, existem métodos para aproximação do cálculo integrais com excelente precisão, e para o caso da distribuição Normal, esses valores já estão tabelados.

Dito isso, vamos aprender/relembrar como calcular as principais probabilidades envolvendo a distribuição Normal.

Definição 2.2

Cálculos com a Distribuição Normal é denominada **Normal Padrão**. Denotamos $Z \sim N(0, 1)$.

Cálculos com a Distribuição Normal



Exemplo 2.3

Seja $Z \sim N(0, 1)$ a Normal Padrão. Calcule:

a - $P(Z < 1)$;

b - $P(Z < 1.5)$;

c - $P(Z > 2)$;

d - $P(Z > 3)$;

e - $P(-1 < Z < 1)$;

f - $P(0.5 < Z < 1.7)$;

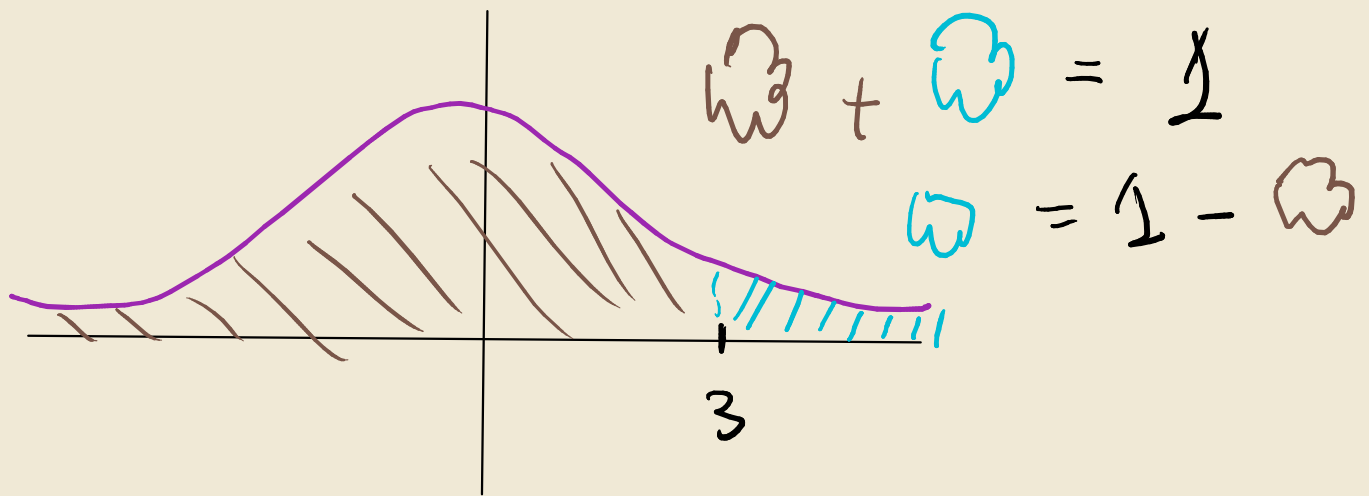
g - $P(Z < -0.7 \text{ ou } Z > 0.7)$;

h - $P(Z < -0.4 \text{ ou } Z > 1)$.

$$a. P(Z < 1) = 0.8413 \approx 0.84$$

$$b. P(Z < 1.5) = 0.9332 \approx 0.93$$

$$c. P(Z > z)$$



$$P(Z > z) = 1 - P(Z < z)$$

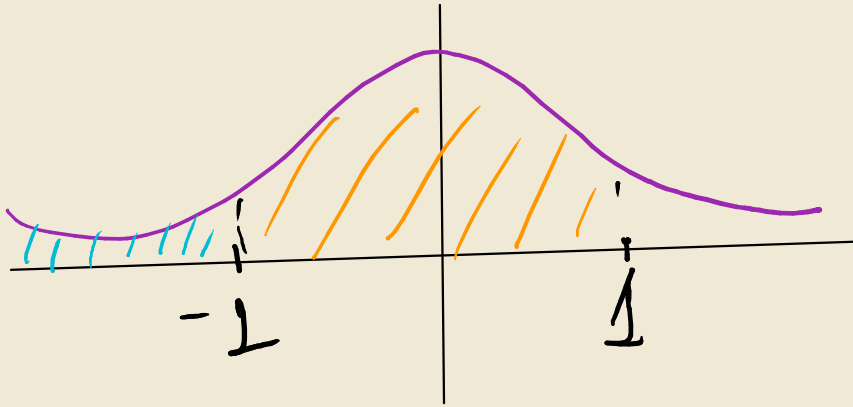
$$= 1 - 0.9772 = 0.0228$$

$$\approx 0.023$$

$$P(Z > 3) = 1 - P(Z < 3)$$

$$= 1 - 0.9987 = 0.0013$$

$$P(-1 < z < 1) =$$



$$P(z < 1) = \text{orange wavy line} + \text{blue cloud shape}$$

$$P(z < -1) = \text{blue cloud shape}$$

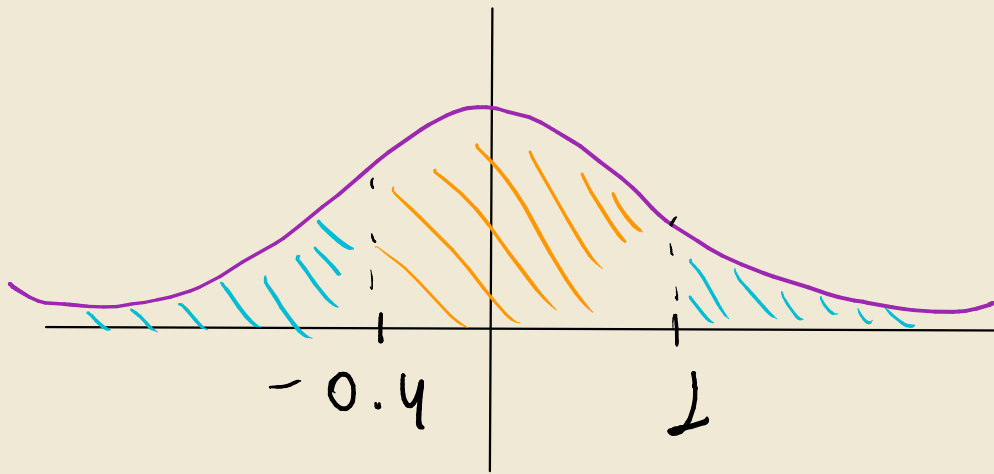
$$P(z < 1) - P(z < -1) = \text{orange wavy line} + \text{blue cloud shape} - \text{blue cloud shape}$$

$$= \text{orange wavy line}$$

$$= 0.8413 - 0.1587 \approx$$

$$0.84 - 0.16 = 0.68$$

$$b- P(Z < -0.4 \text{ or } Z > 1)$$



$$P(Z < -0.4 \text{ or } Z > 1) =$$

$$1 - P(-0.4 < Z < 1) =$$

$$1 - [P(Z < 1) - P(Z < -0.4)] =$$

$$1 - P(Z < 1) + P(Z < -0.4) =$$

$$1 - 0.8413 + 0.3446 \approx$$

$$1 - 0.44 + 0.34 =$$

$$1 - 0.5 = 0.5$$

Exercício 2.1

Seja $Z \sim N(0, 1)$ a Normal Padrão. Calcule:

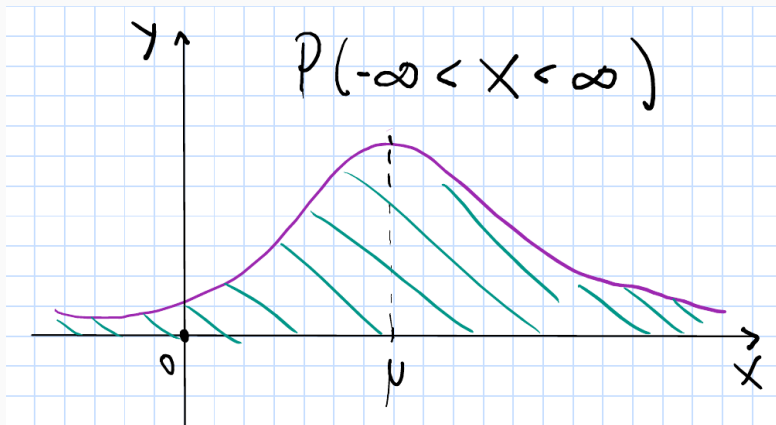
- a - $P(-1 < Z < 1)$;
- b - $P(-2 < Z < 2)$;
- c - $P(-3 < Z < 3)$;
- d - $P(-4 < Z < 4)$.

Vamos usar a Normal Padrão $Z \sim N(0, 1)$ para calcular probabilidades para qualquer normal $X \sim N(\mu, \sigma^2)$.

•

Vamos usar a Normal Padrão $Z \sim N(0, 1)$ para calcular probabilidades para qualquer normal $X \sim N(\mu, \sigma^2)$.

Cálculos com a Distribuição Normal



Para obtermos, a partir da distribuição normal, a distribuição normal padrão ou distribuição normal reduzida, a variável original X é transformada em uma nova variável aleatória Z , com média zero ($\mu = 0$) e variância 1 ($\sigma^2 = 1$):

$$\frac{X - \mu}{\sigma} \sim N(0, 1).$$

Este tipo de transformação, conhecida por **zscore**, é muito utilizada para a padronização de variáveis, pois não altera a forma da distribuição da variável original e gera uma nova variável com média zero e variância 1.

Logo, para calcular $P(X < 0)$ para uma normal $N(\mu, \sigma^2)$ basta aplicar o zscore e calcular

$$P\left(\frac{X - \mu}{\sigma} \leq \frac{0 - \mu}{\sigma}\right)$$

e proceder da mesma maneira que fizemos para o caso $Z \sim N(0, 1)$.

$$X \sim N(8, 36)$$

Diagram illustrating the parameters of the normal distribution $X \sim N(8, 36)$. The mean $\mu = 8$ is indicated by a purple arrow pointing from the number 8 to the letter μ . The variance $\sigma^2 = 36$ is indicated by a purple arrow pointing from the number 36 to the symbol σ^2 .

Vamos ilustrar as contas com a Normal $X \sim N(8, 36)$.

$$\sigma = \sqrt{\sigma^2} = 6$$

Exemplo 2.4

Seja $Z \sim N(8, 36)$ a Normal Padrão. Calcule:

a - $P(X < 0)$;

b - $P(X \leq 12)$;

c - $P(X \leq 20)$;

d - $P(X \geq 2)$;

e - $P(X \geq 5)$;

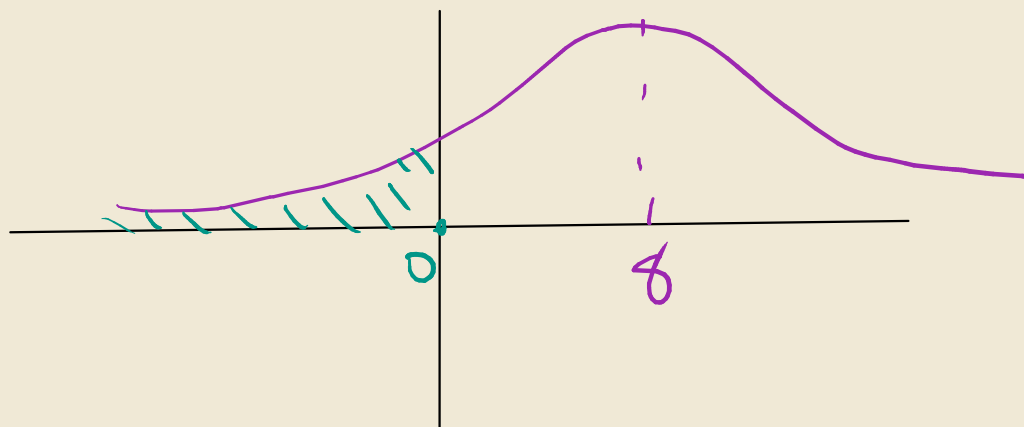
f - $P(6 \leq X \leq 11)$;

g - $P(10 \leq X \leq 25)$;

h - $P(X \leq 3 \text{ ou } X \geq 9)$;

i - $P(X \leq 9 \text{ ou } X \geq 20)$.

$$X \sim N(8, 36)$$



$$P(X < 0) = P\left(\frac{X - \mu}{\sigma} < \frac{0 - \mu}{\sigma}\right)$$

$$= P\left(Z < \frac{0 - 8}{6}\right) =$$

$$P(Z < -1.33) = 0.0918 \approx 0.092$$

$$P(20 \leq X \leq 25)$$



$$P\left(\frac{20 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{25 - \mu}{\sigma}\right) =$$

$$P\left(\frac{20-8}{6} \leq z \leq \frac{25-8}{6}\right) =$$

$$P(0.33 \leq z \leq 2.83) =$$

$$P(z \leq 2.83) - P(z \leq 0.33) =$$

$$0.9977 - 0.6293 \approx$$

$$1 - 0.63 = 0.37$$

Exercício 2.2

Seja $X \sim N(\mu, \sigma^2)$. Calcule:

- a - $P(-\sigma + \mu < Z < \sigma + \mu)$;
- b - $P(-2\sigma + \mu < Z < 2\sigma + \mu)$;
- c - $P(-3\sigma + \mu < Z < 3\sigma + \mu)$;
- d - $P(-4\sigma + \mu < Z < 4\sigma + \mu)$.

Amostragem Aleatória Simples

A amostragem aleatória simples (AAS) é o método mais simples e mais importante para a seleção de uma amostra.

Além de servir como plano próprio, o seu procedimento é usado de modo repetido em procedimento de múltiplos estágios.

A principal caracterização para o uso do plano AAS é a existência de um sistema de referências completo, descrevendo cada uma das unidades elementares.

Deste modo tem-se bem listado o universo

$$\mathcal{U} = \{1, 2, \dots, N\}.$$

O plano é descrito do seguinte modo:

Amostragem Aleatória Simples

- Utilizando algum procedimento aleatório (tabela de números aleatórios, urna, etc) sorteia-se com igual probabilidade um elemento da população \mathcal{U} ;

- repete-se o processo anterior até que sejam sorteados n unidades, tendo sido este número pré-fixado anteriormente;

- caso seja permitido o sorteio de uma unidade mais de uma vez, tem-se o processo AAS com reposição (AASc). Quando o elemento sorteado é removido de \mathcal{U} antes do sorteio próximo, tem-se o plano AAS sem reposição (AASs).

$$U = \{1, 2, 3\}$$

$$AAS_c(z) = \{(11), (12), (13), (21), (22), (23), (31), (32), (33)\}$$

$$AAS_s(z) = \{(12), (13), (21), (23), (31), (32)\}$$

$$P(X=2 \mid X \neq 1)$$

$$P(X=2)$$

Do ponto de vista prático, o plano AASs é muito mais interessante, pois vai de encontro ao princípio intuitivo que "não se ganha mais informação se uma mesma unidade aparece mais de uma vez na amostra".

Por outro lado, o plano AASc, introduz vantagens matemáticas e estatísticas, como a independência entre as unidades sorteadas, que facilita em muito a determinação das propriedades estatísticas dos estimadores das quantidades populacionais de interesse.

Basta observar na maioria dos assuntos tratados em livros de inferência há imposição de que as unidades que fazem parte da amostra sejam independentes.

Também vamos considerar inferências para os seguintes parâmetros de interesse: considere para cada unidade i , uma característica populacional unidimensional de interesse, Y_i , $i \in \mathcal{U}$. Vamos considerar:

- total populacional,

$$\tau = \sum_{i=1}^N Y_i;$$

- média populacional,

$$\mu = \bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i;$$

- variância populacional, representada por

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (Y_i - \mu)^2 \text{ e } S^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \mu)^2.$$

Amostragem Aleatória Simples Com Reposição (AASc)

Amostragem Aleatória Simples Com Reposição (AASc)

A AASc opera da seguinte forma:

- A população está numerada de 1 a N , de acordo com o sistema de referências, ou seja,

$$\mathcal{U} = \{1, 2, \dots, N\}.$$

Amostragem Aleatória Simples Com Reposição (AASc)

- Utilizando um software ou tabela de números aleatórios, sorteia-se, com igual probabilidade, uma das N unidades da população;

Amostragem Aleatória Simples Com Reposição (AASc)

- repõe-se essa unidade na população e sorteia-se um elemento seguinte;

Amostragem Aleatória Simples Com Reposição (AASc)

- repete-se esse procedimento até que n unidades tenham sido sorteadas.

Amostragem Aleatória Simples Com Reposição (AASc)

Agora vamos listar os resultados teóricos envolvendo AASc:

Teorema 4.1

Para o plano amostral AASc, a variável f_i , número de vezes que a unidade i aparece na amostra segue uma distribuição binomial

$$f_i \sim b\left(n, \frac{1}{N}\right).$$

Além disso,

$$E[f_i] = \frac{n}{N}, \text{ Var}[f_i] = \frac{n}{N} \left(1 - \frac{1}{N}\right)$$
$$\text{Cov}[f_i, f_j] = -\frac{n}{N} \text{ para } i \neq j.$$

Teorema 4.2

A estatística $t(\mathbf{s})$, total da amostra, definida por

$$t(\mathbf{s}) = \sum_{i \in \mathbf{s}} Y_i$$

tem, para o plano AASc, as seguintes propriedades:

$$E[t] = n\mu \text{ e } \text{Var}[t] = n\sigma^2.$$

Dos resultados acima, derivam-se estimadores não-viesados para μ e σ^2 , resumidos no seguinte teorema.

$$E[\bar{Y}] = \mu$$

Teorema 4.3

A média amostral

$$\bar{y} = \hat{\mu} = \frac{t(s)}{n} = \frac{1}{n} \sum_{i \in s} Y_i$$

é um estimador não viesado da média populacional μ dentro do plano AASc, e ainda

$$\text{Var}[\bar{y}] = \frac{\sigma^2}{n}.$$

$$E[s^2] = \sigma^2$$

Teorema 4.4

Dentro do plano AASc, a estatística

$$s^2 = \frac{1}{n-1} \sum_{i \in s} (Y_i - \bar{y})^2,$$

é um estimador não viesado da variância populacional σ^2 .

Exemplo 4.5

Considere novamente os dados na planilha "aula-02-exemplo", e considere a variável renda familiar, onde o universo é \mathcal{U} e o parâmetro populacional é $\mathbf{D} = (12, 30, 18)$. Vamos verificar como se comportam \bar{y} e s^2 com relação as funções paramétricas μ e σ^2 de \mathbf{D} .

$$U = \{1, 2, 3\}$$

$$D = (12, 30, 18)$$

1	12
2	30
3	18

Parâmetro (em geral desviante)

$$N = \frac{12 + 30 + 18}{3} = 20$$

$$\sigma^2 = \frac{(12-20)^2 + (30-20)^2 + (18-20)^2}{3}$$

$$= \frac{(-8)^2 + (10)^2 + (-2)^2}{3}$$

$$= \frac{64 + 100 + 4}{3} = \frac{168}{3} = 56$$

$$s^2 = \frac{(12-20)^2 + (30-20)^2 + (18-20)^2}{3-1}$$

$$= \frac{168}{2} = 84$$

$$AAS_c(z)$$

s	(11)	(12)	(13)	(21)	(22)	(23)	(31)	(32)	(33)
\bar{y}	12	21	15	21	30	24	15	24	18
s^2	0	81	217	81	0	36	117	36	0

h	12	15	18	21	24	30
$P(\bar{y}=h)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

h	0	36	81	117
$P(s^2=h)$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{2}{9}$

$$E[\bar{Y}] = \sum \bar{y}_i P(\bar{Y} = \bar{y}_i)$$

$$= 12 \cdot \frac{1}{9} + 15 \cdot \frac{2}{9} + 18 \cdot \frac{1}{9} + 21 \cdot \frac{2}{9} + 24 \cdot \frac{2}{9} + 30 \cdot \frac{1}{9}$$

$$= \frac{12 + 30 + 18 + 42 + 48 + 30}{9} =$$

$$= \frac{42 + 60 + 78}{9} = \frac{102 + 78}{9} = \frac{180}{9} = 20$$

s	(11)	(12)	(13)	(21)	(22)	(23)	(31)	(32)	(33)
\bar{Y}	12	21	15	21	30	24	15	24	18
$\bar{Y} \cdot p_i$	8	1	5	1	10	4	5	4	2

$$\boxed{\mu = 20}$$

↑ erro que cometemos

h	0	36	81	117
$P(S^2=h)$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{2}{9}$

$$E(S^2) = 0 \cdot \frac{3}{9} + 36 \cdot \frac{2}{9} + 81 \cdot \frac{2}{9} + 117 \cdot \frac{2}{9}$$

$$= \frac{72 + 162 + 234}{9} = \frac{468}{9} = 52 = \sigma^2$$

Normalidade e Intervalo de Confiança

Conforme o tamanho da amostra aumenta, as distribuições de \bar{y} e T vão se aproximando da distribuição normal, de acordo com o Teorema do Limite Central (TLC), tanto para o caso da AASc como para a AASs.

Teorema 5.1 (Limite Central)

Para amostras aleatórias simples (X_1, \dots, X_n) , retiradas de uma população com média μ e variância σ^2 finita, a distribuição amostral da média \bar{X} aproxima-se, para n grande, de uma distribuição normal, com média μ e variância σ^2/n .

Então para n suficientemente grande, temos, com relação à AASc, que

$$\frac{\bar{y} - \mu}{\sqrt{\sigma^2/n}} \sim N(0, 1) \text{ e } \frac{T - \tau}{N\sqrt{\sigma^2/n}} \sim N(0, 1),$$

onde $N(0, 1)$ denota uma variável aleatória com distribuição normal com média zero e variância 1.

Os resultados acima possibilitam a obtenção de intervalos de confiança aproximados para \bar{y} e T .

Então com relação à média populacional, temos para n suficientemente grande que

$$P\left(\frac{|\bar{y} - \mu|}{\sqrt{\sigma^2/n}} \leq z_\alpha\right) \simeq 1 - \alpha,$$

onde z_α é a ordenada da $N(0, 1)$ de tal forma que a área na densidade da $N(0, 1)$ no intervalo $(-z_\alpha, z_\alpha)$ é igual a $1 - \alpha$.

Quando σ^2 for desconhecido, iremos substituí-lo por seu estimador não viciado s^2 , que para n grande é bem próximo de σ^2 . Assim, a expressão acima pode ser escrita como

$$P\left(\bar{y} - z_{\alpha}\sqrt{\frac{s^2}{n}} \leq \mu \leq \bar{y} + z_{\alpha}\sqrt{\frac{s^2}{n}}\right) \simeq 1 - \alpha.$$

Disso segue que

$$\left(\bar{y} - z_{\alpha} \sqrt{\frac{s^2}{n}}, \bar{y} + z_{\alpha} \sqrt{\frac{s^2}{n}} \right)$$

é um intervalo de confiança para μ com coeficiente de confiança aproximadamente igual a $1 - \alpha$.

A interpretação usual do intervalo de confiança está baseada no fato de que se forem observadas 100 amostras AAS, e construídos 100 intervalos de confiança baseados nestas amostras, então, aproximadamente $100(1 - \alpha)\%$ dos intervalos devem conter μ .

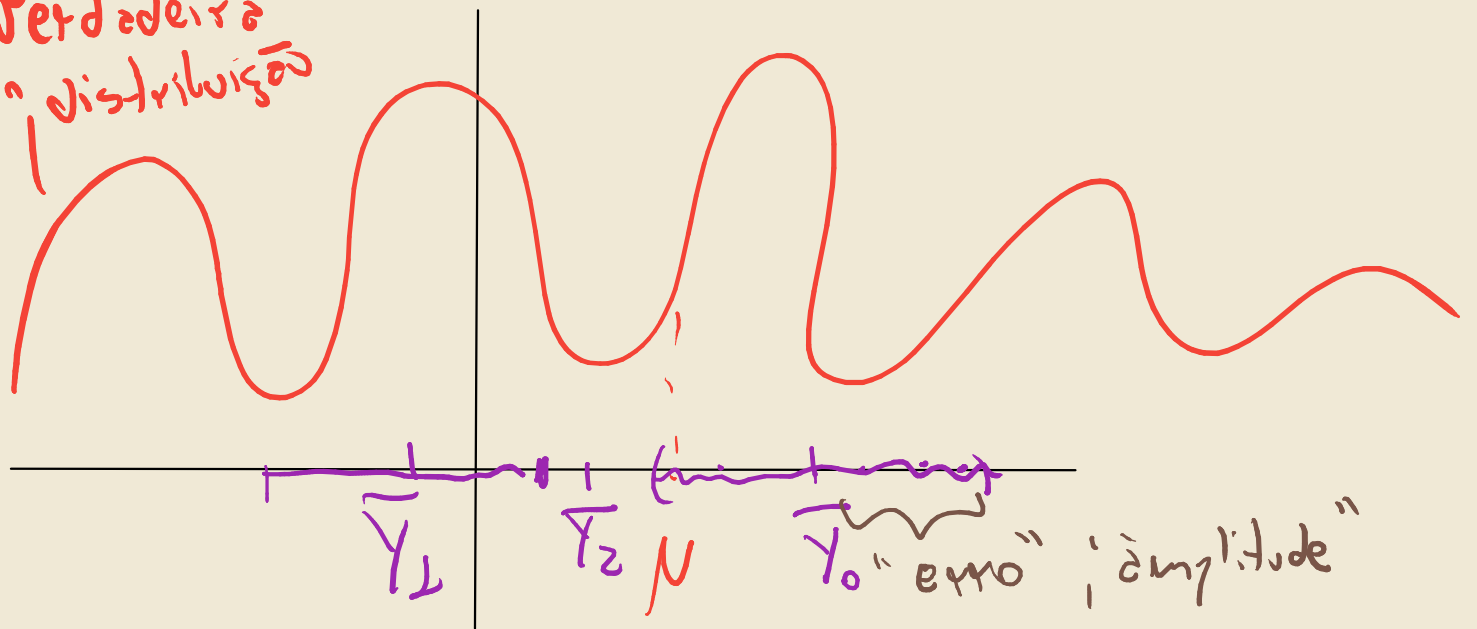


Exemplo 5.2

Uma máquina enche pacotes de café com uma variância igual a 100g. Ela estava regulada para encher os pacotes com 500g, em média. Agora, ela se desregulou, e queremos saber qual a nova média μ . Uma amostra de 25 pacotes apresentou uma média igual a 485g. Vamos construir um intervalo de confiança com 95% de confiança para μ .

$$1 - \alpha$$

verdadeira
distribuição



$$P\left(\frac{|\bar{Y} - \mu|}{\sqrt{\sigma^2/n}} \leq z_\alpha\right) \approx 1 - \alpha$$

$$P(z \leq z_\alpha) \approx 0.95$$

$$z_\alpha \approx 1.65$$

$$\left(\bar{y} - z_{\alpha} \sqrt{\frac{\sigma^2}{n}} ; \bar{y} + z_{\alpha} \sqrt{\frac{\sigma^2}{n}} \right)$$

$$= \left(485 - 1.65 \sqrt{\frac{100}{25}} ; 485 + 1.65 \sqrt{\frac{100}{25}} \right)$$

$$= (485 - 1.65 \cdot 2 ; 485 + 1.65 \cdot 2)$$

$$= (485 - 3.3 ; 485 + 3.3) =$$

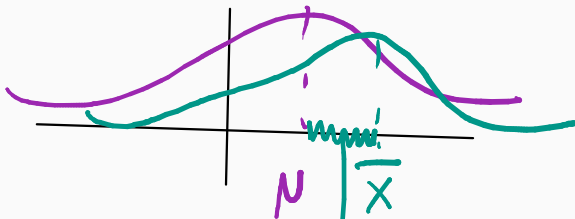
$$= (481.7 ; 488.3)$$

↳ Com 95% de chance

$$\mu \in (481.7 ; 488.3)$$

Determinação do Tamanho da Amostra

Determinação do Tamanho da Amostra



Agora vamos determinar o tamanho da amostra n de tal forma que o estimador obtido tenha um erro máximo de estimação igual a ε , com determinado grau de confiança (probabilidade).

z_{α} ε

Determinação do Tamanho da Amostra

Para a determinação do tamanho da amostra é preciso:

- fixar o erro máximo desejado ε ;
- fixar o grau de confiança z_α ;
- possuir algum conhecimento da variabilidade da população σ^2 .

Os dois primeiros são fixados pelo pesquisador, e quanto ao terceiro, a resposta exige mais trabalho. O uso de pesquisas anteriores ou amostras piloto são os critérios mais utilizados.

De maneira mais específica, o problema consiste em determinar n de modo que

$$P(|\bar{y} - \mu| \leq \varepsilon) \simeq 1 - \alpha.$$

Para n grande, tem-se que

$$P \left(|\bar{y} - \mu| \leq z_{\alpha} \sqrt{\frac{\sigma^2}{n}} \right) \simeq 1 - \alpha.$$

Então para ε fixado, a solução do problema acima consiste em determinar n de tal forma que

$$n \geq \frac{\sigma^2}{(\varepsilon/z_\alpha)^2}.$$

Determinação do Tamanho da Amostra

$$|\bar{x} - \mu| \leq 4$$

Exemplo 6.1

Considere a população de moradores de um condomínio ($N = 540$). Deseja-se estimar a idade média dos condôminos. Com base em pesquisas passadas, pode-se obter a estimativa para σ^2 de 463.32. Suponha que será retirada da população uma amostra segundo AASc. Admitindo que a diferença entre a média amostral e a verdadeira média populacional seja, no máximo, de 4 anos, com um nível de confiança de 95%, determine o tamanho da amostra a ser coletada.

$$n \gg \frac{463.32}{4^2 z_\alpha^2} = \frac{463.32}{4^2 \cdot 1.65^2} =$$

$$P(Z \leq z_\alpha) = 0.95$$

$$z_\alpha = 1.65$$

$$\frac{463.32}{10.89} \approx 42.5$$

$$10.89$$

$$n \gg 43$$

Estimação da Proporção

De maneira geral, em muitas situações, existe interesse em estudar a proporção de elementos em certa população que possuem determinada característica, como ser ou não um item defeituoso, ser ou não eleitor de um determinado partido, etc.

Nesta situação, a cada elemento da população está associada a variável aleatória Y_i da seguinte maneira:

$$Y_i = \begin{cases} 1 & \text{se o indivíduo for portador da característica} \\ 0 & \text{se o indivíduo não for portador da característica} \end{cases}$$

Daí

$$P = \mu = \frac{1}{N} \sum_{i=1}^N Y_i$$

é a proporção de unidades da população que possuem a característica de interesse, e podemos escrever

$$\sigma^2 = P(1 - P).$$

A seguir segue um resumo dos resultados teóricos obtidos para a estimação da proporção:

Teorema 7.1

Um estimador não-viciado de P baseado na AASc é dado por:

$$p = \hat{P} = \bar{y} = \frac{m}{n},$$

com

$$\text{Var}(\hat{P}) = \frac{PQ}{n},$$

sendo $Q = 1 - P$. Além disso, um estimador não-viciado de $\text{Var}(P)$ é

$$\text{var}(p) = \frac{\hat{P}\hat{Q}}{n-1}.$$

Utilizando-se a aproximação normal da binomial, um intervalo de confiança aproximado para P é dado por

$$\left(\hat{P} - z_{\alpha} \sqrt{\frac{\hat{P}\hat{Q}}{n-1}}; \hat{P} + z_{\alpha} \sqrt{\frac{\hat{P}\hat{Q}}{n-1}} \right).$$

Notando-se que o produto PQ (e portanto $\hat{P}\hat{Q}$) é sempre menor que $1/4$, um intervalo de confiança conservador para P é dado por

$$\left(\hat{P} - z_{\alpha} \sqrt{\frac{1}{4(n-1)}}; \hat{P} + z_{\alpha} \sqrt{\frac{1}{4(n-1)}} \right).$$

Como no caso da Média amostral, pode-se considerar o tamanho da amostra n de tal forma que

$$P(|\hat{P} - P| \leq \varepsilon) \cong 1 - \alpha.$$

Pode-se mostrar que o valor de n é dado por

$$n = \frac{PQ}{\varepsilon^2 / z_{\alpha}^2},$$

e para usarmos esta fórmula, é necessário um valor estimado para P .

Uma forma alternativa, que produz um valor conservador para n consiste em utilizar o fato de que $PQ \leq 1/4$. Neste caso, tem-se

$$n \geq \frac{z_{\alpha}^2}{4\varepsilon^2}.$$

$$0.30(1 - 0.3) = 0.7 \times 0.3$$

Exemplo 7.2

Suponha que $p = 30\%$ dos estudantes de uma escola sejam mulheres. Colhemos uma AAS de $n = 10$ estudantes e calculamos

\hat{p} = proporção de mulheres na amostra.

Qual a probabilidade de que \hat{p} difira de p em menos de 0,01?

$$P(|\hat{p} - p| \leq 0.01) =$$

$$P\left(\frac{|\hat{p} - p|}{\sigma/\sqrt{n}} \leq \frac{0.01}{\sigma/\sqrt{n}}\right) =$$

$$P\left(z \leq \frac{0.01}{\sqrt{\frac{0.3 \times 0.7}{10}}}\right) =$$

$$P\left(z \leq \frac{0.01}{\sqrt{0.021}}\right) = P\left(z \leq \frac{0.01}{0.145}\right)$$

$$P(z \leq 0.067) = 0.5239$$

Comentários Finais

Em resumo, na aula de hoje nós lidamos com os aspectos da AASc:

- definição da AASc;
- propriedades das principais estatísticas;
- normalidade e intervalo de confiança;
- tamanho da amostra;
- estimação da proporção.

Nas próximas aulas nós vamos:

- fazer exercícios envolvendo os conceitos da aula de hoje;
- começar a lidar com os aspectos da AASs (Amostragem Aleatória Simples Sem Reposição).

ATIVIDADE PARA ENTREGAR (E COMPOR A NOTA N1)

Resolva em grupos de até 4 integrantes os Exercícios 3.5-3.9.

Referências



