Unidade IV

7 PROBABILIDADES

7.1 Noções de teoria de conjuntos

Quando estudadas as probabilidades, principalmente quando isso é feito a partir de dados empíricos, é preciso analisar algumas propriedades do conjunto estudado. Para isso utilizamos a teoria de conjuntos, a qual discutiremos na presente seção.

7.1.1 Noções gerais

Em primeiro lugar, vamos definir alguns conceitos fundamentais e estabelecer a notação apropriada.

Conjuntos e elementos

Os conceitos de conjunto e de elemento fazem parte daqueles que se chamam de conceitos primitivos, que todos sabemos o que é, mas não há como definir formalmente.

Podemos dizer, no entanto, que conjuntos são coleções de elementos, sendo que elementos de um mesmo conjunto devem ter algo em comum; não faz muito sentido pensar em um conjunto em que um elemento seja uma flor, outro seja um planeta e o terceiro seja uma letra de música. Ou seja, temos sempre conjuntos de algo que podemos definir como sendo de uma mesma classe de elementos: um conjunto de pessoas, um conjunto de países, um conjunto de móveis etc.

Representação

Utilizamos geralmente uma letra maiúscula para nomear o conjunto, e seus elementos são listados entre chaves, como segue:

- $A = \{j, l, n, p\}$

- $B = \{3, 5, 89\}$

- C = {Pedro, João, Nádia}

Note que a lista de elementos pode ser dada numa ordem qualquer, desse modo posso escrever o conjunto A de maneiras diferentes. Por exemplo, se tenho $A = \{j, l, n, p\}$, posso escrever o mesmo conjunto como $A = \{l, j, p, n\}$, visto que em ambos os casos temos os mesmos elementos.

Tomando como exemplo os conjuntos anteriores, podemos dizer que o elemento j pertence ao conjunto A, que o elemento 3 pertence ao conjunto B etc. Podemos ainda dizer que o elemento 5 não pertence ao conjunto A ou que Simone não pertence ao conjunto C. A maneira de escrevermos essas informações em linguagem matemática é a seguinte:

- $j \in A$
- $3 \in B$
- $5 \notin A$
- Simone $\notin C$

Outro ponto importante é a definição do conjunto vazio, indicado pelo fato de não haver nenhum elemento entre as chaves, ou seja, de não haver nenhum elemento no conjunto. Pode parecer estranho definir um conjunto sem elementos, mas quando vamos analisar conjuntos, muitas vezes é importante dizermos que não há nenhum elemento com alguma característica específica. Analogamente, poderíamos pensar que seria estranho termos um número para representar a ausência, a contagem de algo que não há, mas ninguém imaginaria que a existência do número zero não é importante.

Uma notação alternativa para o conjunto vazio é utilizar um símbolo específico, conforme apresentamos a seguir:

$D = \{\ \}$ ou $D = \varnothing$

Subconjuntos

Dizemos que A é subconjunto de B quando todos os elementos que pertencem ao conjunto A também pertencem ao conjunto B.

Vale notar que o conjunto vazio é um subconjunto de qualquer conjunto dado.

Chamamos de conjunto universo, ou simplesmente universo, aquele que reúne todos os elementos que serão considerados no problema.

Exemplo

Para deixar mais claro o conceito de subconjuntos, vamos agora elencar todos os subconjuntos de $A = \{1, 2, 3, 4\}$.

- Conjunto com 4 elementos:

– $A = \{1, 2, 3, 4\}$

- Conjuntos com 3 elementos:

– B = {1, 2, 3}

– C = {1, 2, 4}

– D = {1, 3, 4}

– E = {2, 3, 4}

- Conjuntos com 2 elementos:

– F = {1, 2}

– G = {3, 4}

– H = {1, 4}

– I = {2, 3}

– J = {1, 3}

– K = {2, 4}

- Conjuntos com 1 elemento:

– L = {1}

– M = {2}

– N = {3}

– O = {4}

- Conjunto com 0 elemento:

– P = { }

Vale relembrar que o conjunto A é subconjunto de si próprio e que o conjunto vazio será um subconjunto de A, pois em ambos os casos se atende ao critério de não ter nenhum elemento que não pertença a A.

7.1.2 Operações com conjuntos

Quando temos dois conjuntos, podemos comparar para ver se há elementos em comum ou podemos juntar esses elementos. Esses dois procedimentos se apresentam como sendo as operações de interseção e união de conjuntos, que detalharemos a seguir.

Interseção de conjuntos

A interseção de dois conjuntos A e B será o conjunto que contém os elementos que pertencem a ambos, ao qual chamaremos C.

A linguagem utilizada formalmente para descrever a interseção é descrevê-la por meio do uso da conjunção **e**, ou seja, dizemos que o conjunto C é formado pelos elementos que pertencem a A e B. O símbolo que representa essa operação é um U invertido, ou seja, $\cap$.

Por exemplo, se tivermos A = $\{1, 2, 3, 4\}$ e B = $\{2, 4, 6, 8\}$, a interseção será:

- $C = A \cap B$
- $C = \{2, 4\}$

Pois os elementos 2 e 4 fazem parte dos dois conjuntos.

União de conjuntos

A união de dois conjuntos A e B resulta em um terceiro, ao qual chamaremos C, que tem todos os elementos dos conjuntos A e B. Note que, se houver elementos que pertençam a ambos, eles aparecerão somente uma vez no conjunto C.

A linguagem utilizada formalmente para descrever a interseção é descrevê-la por meio do uso da conjunção **ou**, ou seja, dizemos que o conjunto C é formado pelos elementos que pertencem a A ou B. O símbolo que representa a operação é uma letra U maiúscula.

Num exemplo informal, podemos pensar na seguinte situação: Maria estuda numa escola de idiomas e faz dois cursos, inglês e espanhol. Se a escola resolver fazer uma festa unindo todas as turmas, ela será contada como elemento tanto na turma de inglês como na turma de espanhol, no entanto, isso não faz dela duas pessoas, ela será apenas um elemento no conjunto de alunos da escola.

Tratando agora de um exemplo informal, se tivermos A = $\{1, 2, 3, 4\}$ e B = $\{2, 4, 6, 8\}$, a união será:

- C = A U B
- $C = \{1, 2, 3, 4, 6, 8\}$

Ressalte-se que os elementos 2 e 4, que aparecem nos dois conjuntos, aparecerão somente uma vez no conjunto C, conforme discutido no exemplo da aluna Maria.

Classificação e contagem de elementos

Quando dois conjuntos não têm elementos em comum, dizemos que os conjuntos são disjuntos. Por exemplo, os conjuntos A = $\{1, 4, 6\}$ e o conjunto B = $\{8, 0, 2\}$ são conjuntos disjuntos. Nesses casos, se realizarmos a união dos dois conjuntos, o número de elementos no conjunto final será a soma do número de elementos de cada conjunto.

De uma maneira mais formal, e chamando de n_A, n_B e n_C o número de elementos dos conjuntos A, B e C, respectivamente, teremos:

- $C = A \cup B$
- $C = \{1, 4, 6, 8, 0, 2\}$
- $n_C = n_A + n_B$

Note que essa contagem de elementos numa união de conjuntos, conforme descrita anteriormente, é a origem da operação de adição, podendo muitas vezes até se confundir com ela.

No entanto, se houver elementos comuns aos dois conjuntos, se simplesmente somarmos os números de elementos de A e B, estaremos contando os elementos repetidos duas vezes, retomando o exemplo da Maria, ela seria contada como se fosse duas pessoas em lugar de uma.

Para ilustrar essa situação, retomemos os conjuntos $A = \{1, 2, 3, 4\}$ e $B = \{2, 4, 6, 8\}$.

- $C = A \cup B$, logo $C = \{1, 2, 3, 4, 6, 8\}$
- $D = A \cap B$, logo $D = \{2, 4\}$

Se fôssemos encontrar o número de elementos de C a partir do número de elementos de A e B, somando o número de elementos em cada um, estaríamos contando os elementos 2 e 4 duas vezes. Para evitar essa contagem indevida, teríamos que somar os números de elementos dos dois conjuntos e descontar o número de elementos contados duas vezes. Formalmente, teríamos:

- $n_C = n_A + n_B - n_D$

Ou, de uma maneira mais genérica:

- $n_{A \cup B} = n_A + n_B - n_{A \cap B}$

Chamamos essa formulação de princípio aditivo.

Lembrete

É comum encontrar problemas de contagem de elementos, utilizando a teoria de conjuntos a serem achados, em provas de raciocínio lógico em concursos.

Exemplo

Para ilustrar a questão da contagem, suponhamos que dois amigos, Pedro e Luiz, resolvem fazer uma festa conjunta e cada um deles elabora uma lista de convidados com 20 pessoas. No entanto, eles têm amigos em comum, e isso se reflete no fato de 12 pessoas constarem das duas listas.

Com essas informações e a utilização do princípio aditivo, podemos deduzir o número efetivo de convidados.

Chamando de P o conjunto de convidados de Paulo e de L o conjunto de convidados de Luiz, temos:

- Convidados de Paulo: $n_P = 20$
- Convidados de Luiz: $n_L = 20$
- Convidados em comum: $n_{P \cap L} = 12$
- Convidados da festa: $n_{P \cup L} = n_P + n_L - n_{P \cap L}$
- Convidados da festa: $n_{P \cup L} = 20 + 20 - 12$
- Convidados da festa: $n_{P \cup L} = 28$

Ou seja, embora cada um tenha 20 convidados, será 28 o número total de pessoas chamadas a participar da festa.

7.1.3 Diagramas de Venn

Os diagramas de Venn foram desenvolvidos com o objetivo de tornar fácil a visualização de conjuntos e subconjuntos. Nesta seção vamos mostrar como se constroem esses diagramas e como podemos utilizar essa ferramenta como um recurso auxiliar na análise de probabilidades.

Montagem de diagramas com os elementos

Nesses diagramas, associa-se uma área para cada conjunto, na qual ficam seus elementos. Quando um elemento pertence a mais de um conjunto, haverá uma área comum aos dois conjuntos, de modo que exista um lugar para colocar esse elemento.

No diagrama apresentado no gráfico 9, temos uma representação referente aos conjuntos utilizados num exemplo da seção 2: $A = \{1, 2, 3, 4\}$ e $B = \{2, 4, 6, 8\}$.

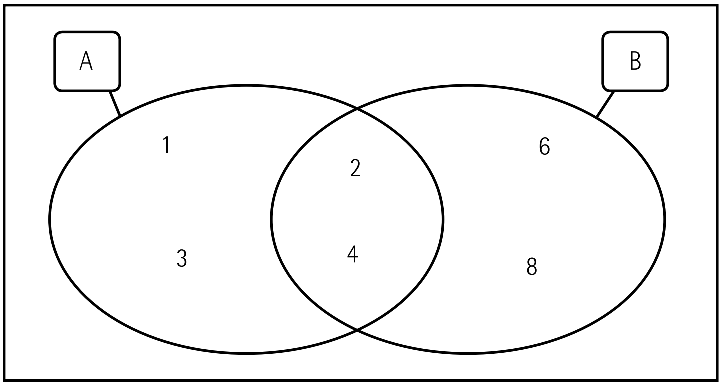


Gráfico 9 – Representação de dois conjuntos com interseção.

Assim:

- Na região central, a qual pertence tanto à elipse do conjunto A quanto à do conjunto B, são colocados os elementos comuns aos dois conjuntos: 2 e 4.

- Na região que pertence à elipse do conjunto A, mas não à elipse do conjunto B, foram colocados os elementos que pertencem a A, mas não a B: 1 e 3.

- Na região que pertence à elipse do conjunto B, mas não à elipse do conjunto A, foram colocados os elementos que pertencem a B, mas não a A: 6 e 8.

Desse modo, a elipse A tem os quatro elementos do conjunto A e a elipse B tem os quatro elementos do conjunto B, sendo que todos os elementos dos conjuntos foram representados de modo tal que podemos, apenas com o uso do diagrama, saber rapidamente quais são os elementos de cada conjunto e quais são ou não comuns a ambos.

Suponhamos agora que nosso conjunto universo englobe todos os números inteiros entre 1 e 9. Nesse caso, poderíamos representar essa situação colocando os elementos restantes dentro do quadrado do diagrama, mas fora das elipses, como vemos no gráfico 10.

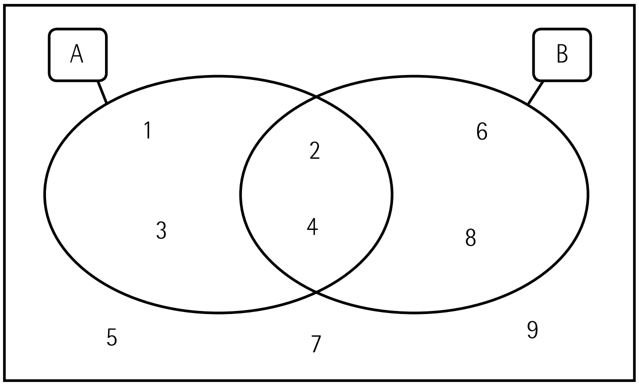


Gráfico 10 – Representação gráfica de dois conjuntos com interseção pertencentes a um conjunto universo com elementos que não pertencem a A ou a B.

Veremos agora alguns exemplos adicionais envolvendo situações diversas da descrita anteriormente, inclusive com representação de três conjuntos.

Observando o gráfico 11, podemos ver que os conjuntos A e B não possuem elemento comum, logo, não há necessidade de haver uma região em comum para os dois conjuntos.

No diagrama do gráfico 11, indicamos o conjunto universo explicitamente, mas tal indicação nem sempre é necessária, ele pode o mesmo ser implicitamente denotado, conforme vimos no gráfico 10.

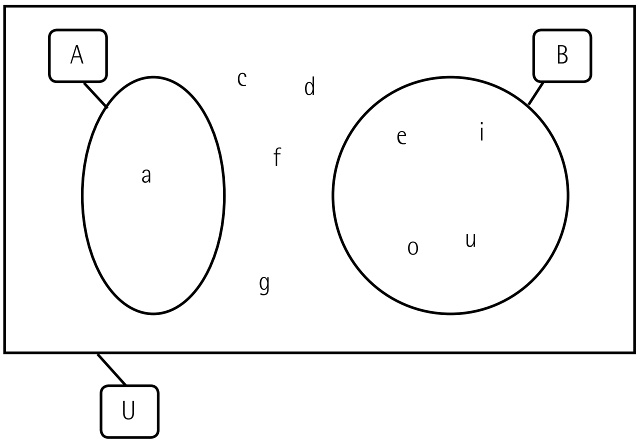


Gráfico 11 – Representação de dois conjuntos sem interseção e conjunto universo indicado explicitamente.

Outra situação de interesse que fica bastante fácil de visualizar é o caso em que temos um conjunto A cujos elementos pertencem também a um conjunto B maior, ou seja, quando A está contido em B. Essa situação está descrita no gráfico 12. Note que o conjunto B tem 3 elementos, todos que estão em sua elipse, ou seja, $B = \{a, b, c\}$.

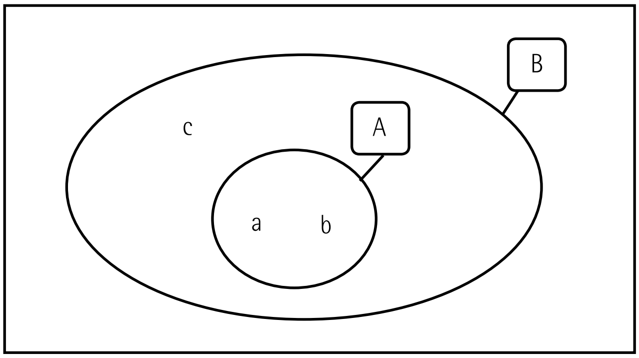


Gráfico 12 – Representação de um conjunto contido em outro.

Tão importante quanto saber montar um diagrama de Venn é saber extrair as informações apresentadas. Para desenvolver tal habilidade, vamos agora efetuar a leitura detalhada de um diagrama com grau de complexidade maior do que os anteriores.

O gráfico 13 traz um diagrama de Venn em que temos três conjuntos, A, B e C, pertencentes ao conjunto universo das letras de nosso alfabeto. Como há diversas regiões representadas, vamos comentar cada uma delas. Desse modo, faremos a leitura detalhada do diagrama.

- A região central, que pertence aos três círculos, traz os elementos *a* e *b*. Portanto, esses são os elementos que pertencem aos três conjuntos A, B e C.
- Na região comum a A e B, mas externa ao círculo C, temos os elementos *d* e *e*. Portanto, esses são os elementos que pertencem aos conjuntos A e B, mas não pertencem a C.
- Na região comum a B e C, mas externa ao círculo A, temos os elementos *f* e *g*. Portanto, esses são os elementos que pertencem aos conjuntos B e C, mas não pertencem a A.
- Na região comum a A e C, mas externa ao círculo B, temos o elemento *c*. Portanto, esse elemento pertence aos conjuntos A e C, mas não pertence a B.
- Na região que pertence ao círculo de A, mas não aos demais círculos, temos os elementos *h, i, j*.
- Na região que pertence ao círculo de B, mas não aos demais, temos o elemento *k*.
- Na região que pertence ao círculo de C, mas não aos demais, temos os elementos *m, n, o, p, q, r, s* e *t*.
- Os elementos representados dentro do quadrado que não estão dentro de nenhum dos círculos pertencem ao universo considerado, mas não pertencem aos conjuntos A, B ou C.

Por fim, vamos especificar cada um dos conjuntos A, B e C. A leitura do diagrama nos mostra que: A = {a, b, c, d, e, h, i, j}, B = {a, b, d, e, f, g, k} e C = {a, b, c, f, g, m, n, o, p, q, r, s, t}.

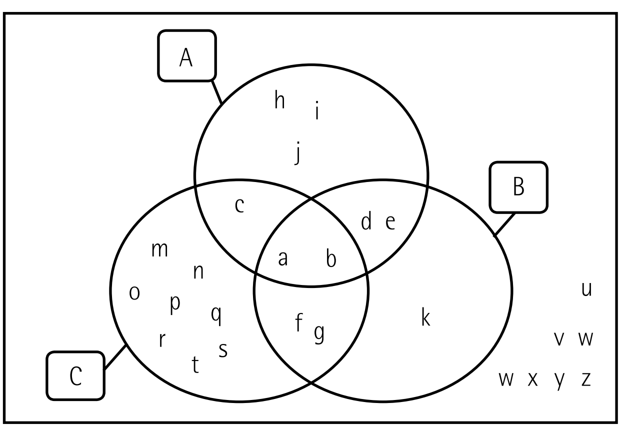


Gráfico 13 – Representação de três conjuntos sem interseção.

Muito embora a leitura de um diagrama como esse não seja conceitualmente complicada, quando temos um número grande de elementos no conjunto universo, a construção e a leitura de tais diagramas começa a perder sua função prática, por excesso de elementos. Além disso, na maioria das vezes a informação de maior interesse é o número de elementos em cada região, sem que haja necessidade de sabermos quais são exatamente os elementos em cada uma delas. Isso é particularmente verdade quando o interesse é calcular probabilidades. Assim, é usual utilizarmos esses diagramas colocando em cada região a contagem de elementos em lugar dos elementos em si, o que faremos na próxima seção.

Diagramas com contagem dos elementos

Vamos agora reapresentar os diagramas de Venn vistos até o momento com a contagem de elementos referente a cada região e vamos ver também como utilizar esses diagramas para extrair informações sobre os conjuntos.

Dada a importância desta leitura e sabendo que a compreensão é às vezes difícil para muitos alunos, vamos apresentar um grande número de exemplos com explicações passo a passo.

No gráfico 14, temos a contagem dos elementos que foram apresentados no gráfico 9. Ali cada campo interno às elipses tem dois elementos.

É importante notar que os números colocados no diagrama de contagem não são mais os elementos do conjunto, mas sim o número de elementos alocados em cada região. Para que isso fique claro, observe novamente o diagrama do gráfico 9.

Ou seja:

- A interseção de A e B tem 2 elementos: $\{2, 4\}$.
- O conjunto A tem 2 elementos que não pertencem ao conjunto B: $\{1, 3\}$.
- O conjunto B tem 2 elementos que não pertencem ao conjunto A: $\{6, 8\}$.
- Todos os elementos do conjunto universo pertencem ao conjunto A ou ao conjunto B.

A partir disso, podemos realizar a contagem de elementos de cada conjunto:

- O conjunto A tem os 2 elementos que não pertencem ao conjunto B mais 2 elementos que pertencem a ele somente, logo $n_A = 4$.
- O conjunto B tem os 2 elementos que não pertencem ao conjunto A mais 2 elementos que pertencem a ele somente, logo $n_B = 4$.
- O conjunto universo tem os 2 elementos que pertencem somente ao conjunto A, 2 elementos que pertencem somente ao conjunto B e mais 2 elementos que pertencem aos dois, logo $n_{tot} = 6$.

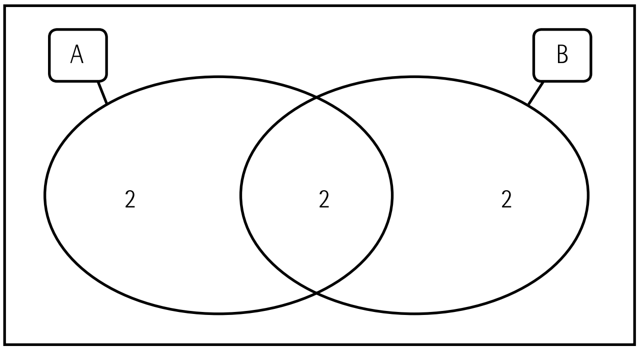


Gráfico 14 – Representação com a contagem de elementos dos conjuntos dados no gráfico 9.

Fazendo agora o mesmo processo para o gráfico 10, construímos o diagrama com a contagem de elementos do gráfico 15.

Na comparação com o gráfico 14, temos como única diferença o fato de haver 2 elementos que são do universo considerado, mas não fazem parte de A nem de B. Logo, tudo o que foi dito para A e B quando lemos o gráfico 14 permanece válido para o gráfico 15.

A alteração se dá na contagem do número de elementos do universo, que é acrescido de 3 elementos. Assim, temos:

- O conjunto A tem os 4 elementos que estão dentro de sua região, ou seja: $n_A = 4$.
- O conjunto B tem os 4 elementos que estão dentro de sua região, ou seja: $n_B = 4$.
- O conjunto universo tem os 6 elementos que pertencem à união dos conjuntos A e B mais os 3 elementos que pertencem somente a ele. Portanto: $n_{tot} = 9$.

Na prática, para obtermos os valores acima, apenas somamos os números que aparecem dentro da região de interesse.

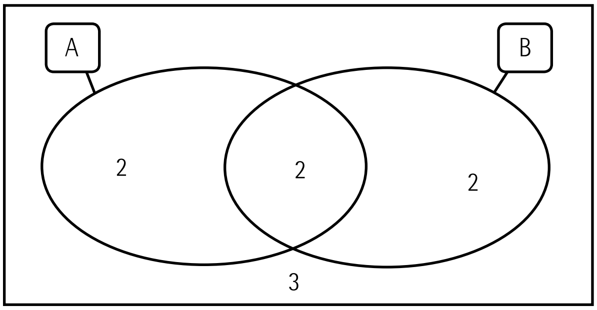


Gráfico 15 – Representação com a contagem de elementos dos conjuntos dados no gráfico 10.

No gráfico 16 colocamos as contagens referentes ao gráfico 11. Nesse caso, como os conjuntos A e B são disjuntos, a determinação de seu número de elementos consiste simplesmente em ler o número de elementos na região pertencente a cada um. O número de elementos do conjunto universo será a soma dos números escritos no diagrama.

Assim, teremos: $n_A = 1$, $n_B = 4$ e $n_{tot} = 9$.

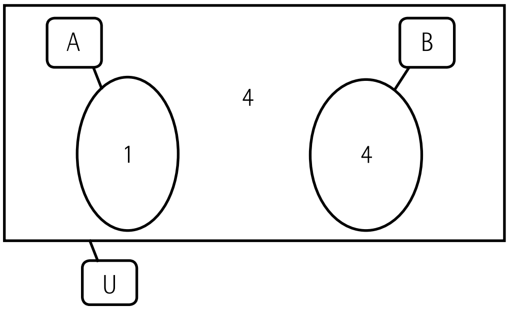


Gráfico 16 – Representação com a contagem de elementos dos conjuntos dados no gráfico 11.

Para o diagrama apresentado no gráfico 17, a leitura nos dará os valores: $n_A = 2$, $n_B = 3$ e $n_{tot} = 3$, visto que todos os elementos do conjunto universo fazem parte do conjunto B.

Note ainda que o número zero colocado no diagrama tem o papel de indicar que aquela região não tem nenhum elemento. Em princípio, poderíamos deixar a região em branco, mas, na prática, quando vamos preencher diagramas é mais indicado que o valor zero seja indicado, pois a ausência de contagem pode fazer crer que, durante a análise, se tenha esquecido de considerar aquela região.

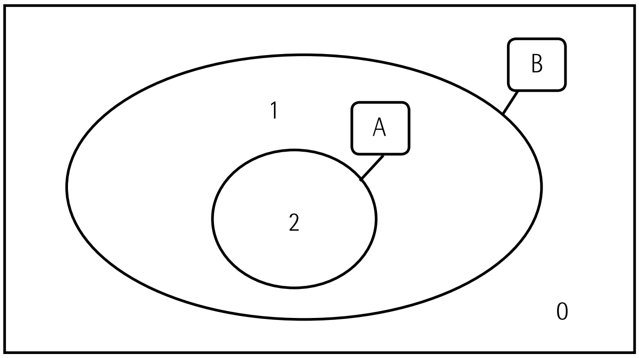


Gráfico 17 – Representação com a contagem de elementos dos conjuntos dados no gráfico 12.

O caso com três conjuntos, A, B e C, apresentado no gráfico 18 a partir do diagrama de elementos do gráfico 13 exige uma leitura mais cuidadosa, logo, vamos efetuá-la passo a passo.

Em primeiro lugar, notamos que em cada região está colocado o número de elementos correspondente, do mesmo modo que nos demais diagramas já apresentados nesta seção. O aluno que tem dificuldade na construção e leitura de diagramas de Venn deve buscar construir o diagrama a partir do original, de modo a compreender de maneira apropriada o significado de cada região dele.

Fazendo a leitura individual de cada região, teremos:

- Elementos que não pertencem a A, B ou C: 7.
- Elementos pertencentes a A, B e C: 2.
- Elementos pertencentes a A e B, mas não a C: 2.
- Elementos pertencentes a B e C, mas não a A: 2.
- Elementos pertencentes a A e C, mas não a B: 1.
- Elementos pertencentes somente a A: 3.
- Elementos pertencentes somente a B: 1.
- Elementos pertencentes somente a C: 8.

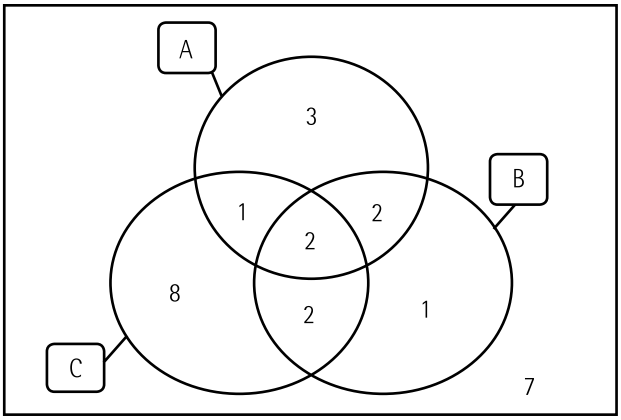


Gráfico 18 – Representação com a contagem de elementos dos conjuntos dados no gráfico 13.

Feita a leitura dos campos individualmente, passamos agora a ver situações que englobam mais de um campo e que, eventualmente, geram dúvidas quando vamos utilizar esses diagramas.

Nos exemplos que seguem, vamos utilizar o artifício de deixar em branco somente os campos que contêm elementos que serão contados na situação descrita, facilitando a compreensão do uso do diagrama.

Vamos agora ver como determinar o número de elementos nos conjuntos A, B e C.

Se quisermos saber quantos elementos pertencem ao conjunto A, temos que somar os números que estão dentro do círculo A, conforme ilustramos no gráfico 19. A quantidade de elementos no conjunto A será a soma de todos os elementos que estão dentro de seu círculo, $n_A = 3 + 2 + 2 + 1$, logo, $n_A = 8$, o que pode ser facilmente confirmado se olharmos os elementos de conjunto A colocados no gráfico 13.

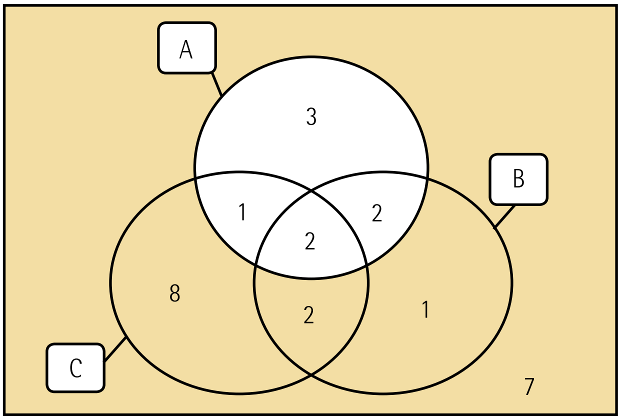


Gráfico 19 – Representação com a contagem de elementos dos conjuntos dados no gráfico 13, enfatizando o conjunto A.

Se procedermos da mesma maneira, somando os elementos de cada campo do respectivo círculo, encontraremos $n_B = 7$ e $n_C = 13$.

Muitas vezes, é de interesse verificar algumas situações que envolvem algumas características que não estão dadas em um campo ou no círculo completo. Por exemplo, podemos querer o número de elementos que pertence a cada um dos conjuntos. Nesse caso, teremos que somar os elementos que pertencem somente a A com os pertencentes apenas a B e os que são apenas de C, obtendo n = 3 + 1 + 8 = 12. Essa situação está ilustrada no gráfico 20.

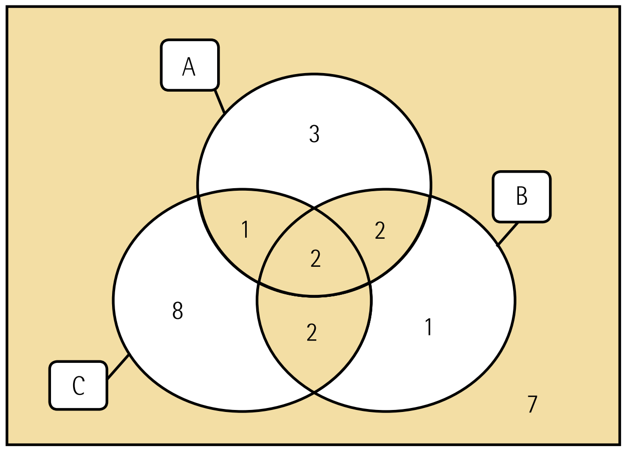


Gráfico 20 – Representação com a contagem de elementos dos conjuntos dados no gráfico 13, enfatizando os elementos que pertencem a cada um dos conjuntos A, B e C.

Outra situação de interesse é saber quantos elementos pertencem a B e C, ou seja, à interseção de C e B. Temos essa situação ilustrada no gráfico 21.

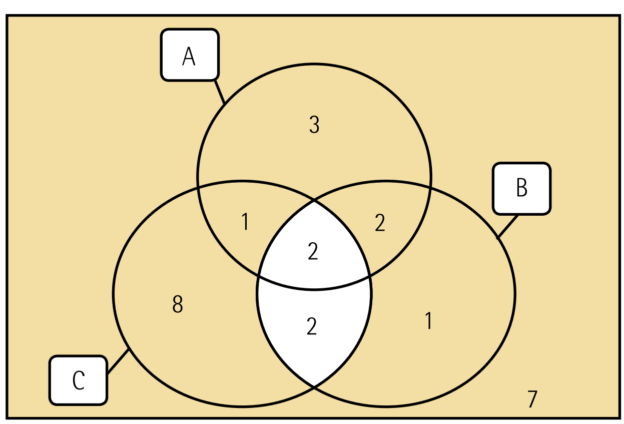


Gráfico 21 – Representação com a contagem de elementos dos conjuntos dados no gráfico 13, enfatizando os elementos que pertencem a B e C.

Um último exemplo de interesse é o caso em que a característica que se busca é definida a partir da exclusão. Por exemplo, quero saber quantos são os elementos que não pertencem ao conjunto B. Procedemos então como se retirássemos o círculo B de dentro da figura e contássemos os elementos restantes, conforme ilustramos no gráfico 22. Teremos assim o número em questão, que é a soma de todos os que estão nos campos não sombreados, resultando em um total de 19 elementos.

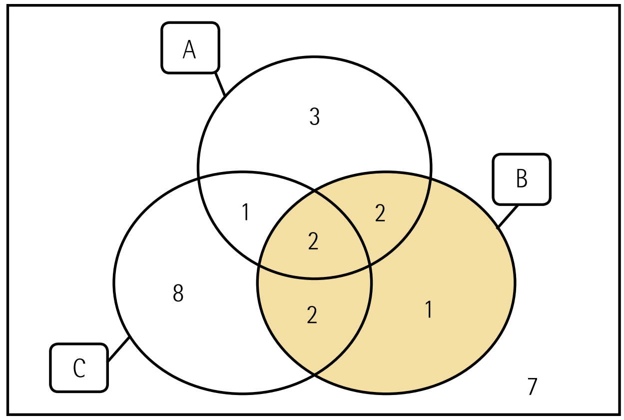


Gráfico 22 – Representação com a contagem de elementos dos conjuntos dados no gráfico 13, enfatizando os elementos que não pertencem a B.

7.1.4 Aplicação

Visto como utilizar os diagramas de Venn para contagem de elementos, passaremos a ver algumas aplicações mais realistas com o uso das informações para calcular probabilidades.

Convidados para festa conjunta de dois amigos

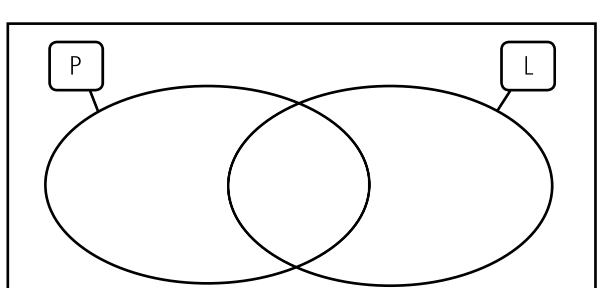
Para ilustrar a montagem do diagrama e começarmos a calcular as probabilidades a partir das informações ali expressas, começaremos com uma variante da situação já discutida anteriormente.

Suponhamos que, em outra festa conjunta, participam 45 convidados de Paulo e 35 convidados de Luiz, havendo 15 pessoas que constam das duas listas. Além disso, acompanhando os convidados, vão à festa 5 pessoas que nenhum deles conhece.

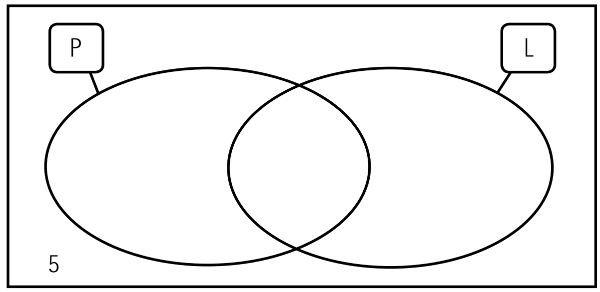
Em primeiro lugar, colocamos as informações em uma tabela para organizá-las de modo a termos mais clareza. Novamente, chamaremos de P e L os conjuntos de convidados de Pedro e Luiz.

P	L	P e L	Nenhum
45	35	15	5

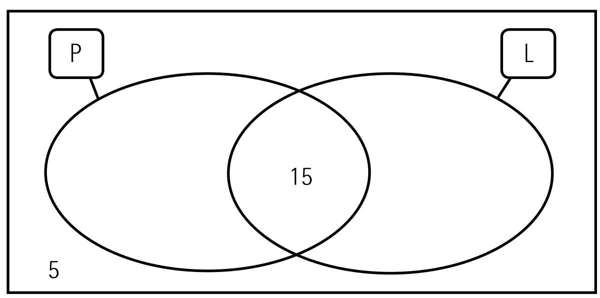
Comecemos agora a organizar essas informações no diagrama. Em primeiro lugar, como há dois conjuntos L e P de maior interesse, e que têm interseção, utilizaremos o diagrama com 2 elipses, como segue:



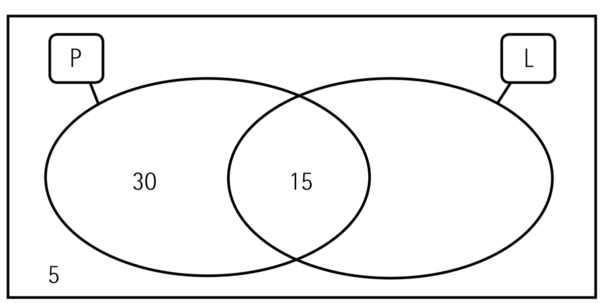
O número mais fácil de colocarmos no diagrama é aquele referente aos acompanhantes, pois ficarão fora das elipses:



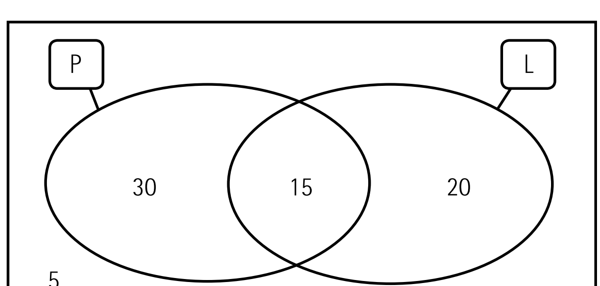
O próximo número a ser colocado é o de convidados de ambos os anfitriões na região pertencente aos dois conjuntos P e L:



Sabemos que são 45 os convidados de Paulo, mas 15 já estão no diagrama, logo, faltam 30, que são convidados somente de Paulo:



Por fim, temos os convidados somente de Luiz, que são 35 menos os 15 convidados em comum, que já estão representados no diagrama. Logo, teremos como representação diagramática com todos os dados o gráfico a seguir:



Tendo essas informações, podemos calcular o número total de participantes da festa sem o risco de contar nenhuma pessoa duas vezes.

$n_{tot} = 30 + 15 + 20 + 5$

$n_{tot} = 70$

De posse dessas informações, passamos a calcular algumas probabilidades:

• Qual a probabilidade de um convidado da festa ser amigo de Luiz?

– Para calcular essa probabilidade P(L), devemos dividir o número de convidados de Luiz (n_L) pelo número total de convidados (n_{tot}).

– Temos $n_L = 35$ e $n_{tot} = 70$, logo, P(L) = 35/70.

– P(L) = 0,5

– Se houver preferência pela notação percentual, P(L) = 50%.

• Qual a probabilidade de um convidado da festa ser amigo de ambos?

– Para calcular P(ambos), devemos dividir o número de convidados de ambos (n_{ambos}) pelo número total de convidados (n_{tot}).

– Temos então $n_{ambos} = 15$ e $n_{tot} = 70$, logo, P(L) = 15/70.

– P(ambos) = 0,21, aproximadamente.

– Em notação percentual, P(ambos) = 21%.

• Qual a probabilidade de um convidado da festa não ser amigo de Pedro?

– Para calcular P(não P), devemos dividir o número de convidados que não faz parte do conjunto P (n_{nP}) pelo número total de convidados (n_{tot}).

– Para calcular n_{nP}, temos duas possibilidades. A primeira é subtrair do total o número de amigos de Pedro, assim: $n_{nP} = 70 - 45 = 25$.

– A segunda possibilidade é somar os números dos campos que não fazem parte do conjunto P, ou seja, $n_{nP} = 20 + 5 = 25$.

– A probabilidade será P(nP) = n_{nP}/n_{tot}, logo, P(L) = 25/70.

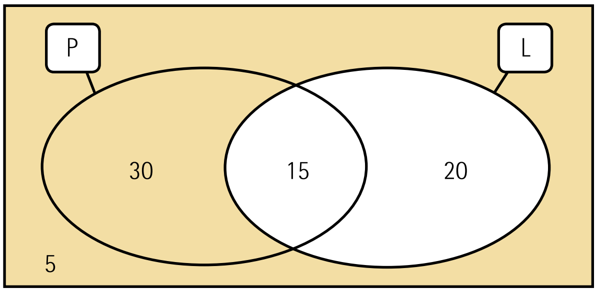
– P(nP) = 0,36, aproximadamente.

– Em notação percentual, P(nP) = 36%.

Nos exemplos anteriores, calculamos probabilidades considerando todos os participantes da festa. Agora, veremos um exemplo em que consideramos um universo mais restrito.

- Qual a probabilidade de um convidado de Luiz ser também amigo de Pedro?

– Para calcular essa probabilidade, temos como universo não mais a totalidade dos convidados, mas apenas aqueles que são amigos de Luiz. Esquematicamente, é como se passássemos a utilizar um diagrama apenas com os elementos pertencentes a L:



– Temos, então, um novo total de eventos possíveis, nesse caso, 35.

– Além disso, só vamos considerar os amigos de Pedro que também são amigos de Luiz e o número de amigos de Pedro nesse universo mais reduzido é de 15 pessoas.

– Assim sendo, essa probabilidade será P = 15/35.

– P(L) = 0,43

– Se houver preferência pela notação percentual, P(L) = 43%.

Clientes de uma empresa

Vamos agora utilizar o mesmo tipo de raciocínio para discutir um exemplo um pouco mais complexo.

Suponhamos que uma empresa que desenvolve programas computacionais tenha três produtos mais importantes, aos quais chamaremos A, B e C.

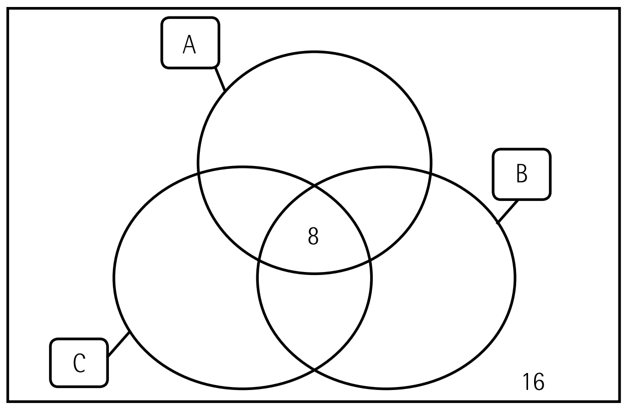
Dos diversos clientes, temos alguns que utilizam apenas um dos programas, alguns utilizam dois deles, outros utilizam os três e há clientes que utilizam somente outros produtos da empresa, conforme colocado no quadro a seguir:

A	B	C	A e B	B e C	A e C	A, B e C	Outros
35	36	36	17	15	16	8	16

Nosso interesse aqui é realizar cálculos de probabilidade referentes a esse universo, logo, é necessário que tenhamos informações relevantes que não estão colocadas no quadro diretamente. Entre outras, precisamos saber quantos clientes a firma tem no total.

Para encontrar esse valor, o primeiro impulso seria somar os números apresentados na tabela, no entanto, a exemplo do que vimos na seção anterior, isso faria com que contássemos um mesmo cliente mais de uma vez. Assim, vamos utilizar um procedimento análogo ao do exemplo da festa organizando essas informações num diagrama de Venn com 3 círculos, cada um deles para um dos programas.

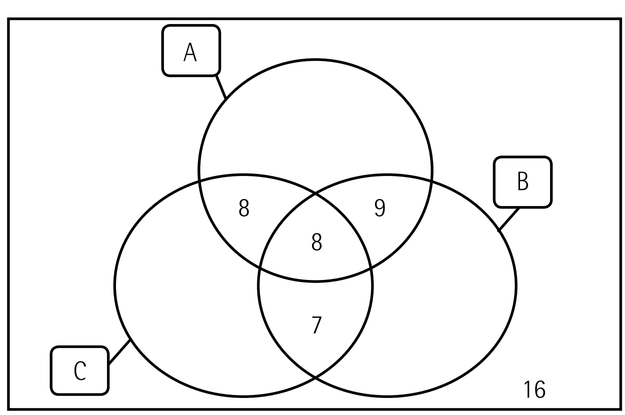
A exemplo do que fizemos anteriormente, o primeiro valor a colocar é o número de clientes que não usa os programas A, B ou C. O outro número que podemos colocar também imediatamente é o dos clientes que utilizam os 3. Desse modo, teremos:



Passamos agora ao preenchimento dos campos em que há usuários de 2 programas. Para isso, temos que descontar do valor dado no quadro os 8 usuários que utilizam os 3 programas e já foram alocados no diagrama.

- 17 usam A e B. Destes, 8 utilizam também C. Logo, 9 utilizam somente A e B.
- 15 usam B e C. Destes, 8 utilizam também A. Logo, 7 utilizam somente B e C.
- 16 usam A e C. Destes, 8 utilizam também C. Logo, 8 utilizam somente A e B.

Colocando essas informações no diagrama, temos:

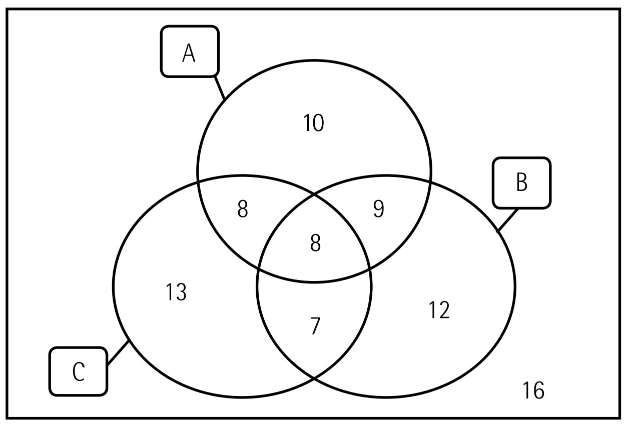


Por fim, vamos ver agora quantos elementos há que utilizam somente um dos programas.

- O programa A tem 35 usuários, dos quais já estão representados 25. Note que no diagrama anterior já há 25 elementos representados dentro do círculo A. Portanto, há 10 clientes que utilizam somente o programa A.

- O programa B tem 36 usuários, dos quais já estão representados 24 dentro do círculo correspondente. Portanto, há 12 clientes que utilizam somente o programa B.
- O programa C tem também 36 usuários, dos quais 23 já estão representados dentro do círculo correspondente. Portanto, há 13 clientes que utilizam somente o programa C.

Com isso, temos agora o diagrama completo:



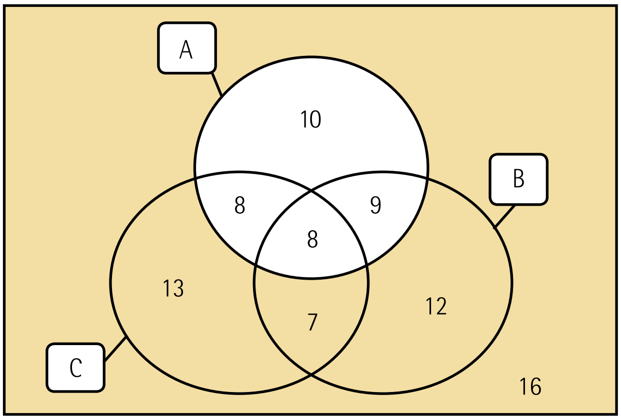
Para sabermos o número total de clientes, basta somar todos os números dados no diagrama e teremos $n_{tot} = 80$. Note que essa soma inclui os 64 usuários de A, B ou C e também os 16 que utilizam apenas outros produtos, mas também são clientes da empresa.

De posse de todas as informações, podemos agora utilizá-las na determinação de probabilidades. Vejamos alguns exemplos.

- Qual a probabilidade de um cliente utilizar o programa A?
  - Eventos favoráveis a A: 35.
  - Eventos possíveis: 80.
  - Probabilidade: P = 35/80.
  - Probabilidade: P = 0,4375 ou P = 43,75%.
- Qual a probabilidade de um cliente utilizar somente o programa A?
  - Eventos favoráveis a somente A: 10.
  - Eventos possíveis: 80.
  - Probabilidade: P = 10/80.
  - Probabilidade: P = 0,125 ou P = 12,5%.

- Qual a probabilidade de um cliente utilizar dois dos programas principais?
  - Eventos favoráveis a usar os dois principais: 9 + 7 + 8 = 24.
  - Eventos possíveis: 80.
  - Probabilidade: P = 24/80.
  - Probabilidade: P = 0,3 ou P = 30%.
- Qual a probabilidade de um cliente não utilizar nenhum dos programas principais?
  - Eventos favoráveis a não usar os principais: 16.
  - Eventos possíveis: 80.
  - Probabilidade: P = 16/80.
  - Probabilidade: P = 0,2 ou P = 20%.
- Qual a probabilidade de um cliente não utilizar o programa B ou C?
  - Eventos favoráveis a não usar B ou C (serão os que usam apenas A somados aos que não usam nenhum dos três): 12 + 16 = 28.
  - Eventos possíveis: 80.
  - Probabilidade: P = 28/80.
  - Probabilidade: P = 0,35 ou P = 35%.

Os exemplos anteriores consideraram sempre como conjunto universo todos os clientes da empresa, mas, assim como fizemos na seção anterior, é possível analisar situações em que haja restrição do conjunto. Por exemplo, se quisermos saber apenas como se comportam os clientes do programa A com relação à utilização dos outros. Nesse caso, o universo considerado será apenas o conjunto representado no círculo A:



- Qual a probabilidade de um usuário de A utilizar também B?
  - Eventos favoráveis a usar B: 9 + 8 = 17 (somente os usuários de B que pertencem ao círculo A).
  - Eventos possíveis: 35.
  - Probabilidade: P = 17/35.
  - Probabilidade: P = 0,4857 ou P = 48,57%.
- Qual a probabilidade de um usuário de A não utilizar B ou C?
  - Eventos favoráveis a não usar B ou C: 10 (os usuários de A somente).
  - Eventos possíveis: 35.
  - Probabilidade: P = 10/35.
  - Probabilidade: P = 0,2857 ou P = 28,57%.
- Qual a probabilidade de um usuário de A não utilizar C?
  - Eventos favoráveis a não usar B: 9 (os usuários de B somente).
  - Eventos possíveis: 35.
  - Probabilidade: P = 9/35.
  - Probabilidade: P = 0,2571 ou P = 25,71%.

7.2 Propriedades importantes

Agora que já há mais familiaridade com o conceito de probabilidade e com a maneira de calculá-la, vamos formalizar algumas propriedades importantes.

7.2.1 Eventos mutuamente exclusivos

Chamamos de eventos mutuamente exclusivos ou excludentes, aqueles cujo acontecimento implica na não ocorrência do outro. Ou seja, se dois eventos A e B são mutuamente exclusivos, quando A ocorre é necessário que B não ocorra.

Podemos citar como exemplos: face cara e face coroa, pois quando temos um resultado, certamente não teremos o outro na mesma jogada; chover e não chover; viajar e ficar em casa; obter as diferentes faces num jogo de dado etc.

Nos casos citados, a probabilidade de acontecer um ou outro evento será a soma das probabilidades de cada evento.

P(A ou B) = P(A) + P(B)

Jogo de dado

Num jogo de dado, os eventos são mutuamente exclusivos, pois a obtenção de uma face implica em não termos obtido nenhuma das outras, logo, podemos utilizar a propriedade anterior para obter a probabilidade de tirar um número par.

• Qual a probabilidade de o resultado ser par?

– Faces pares: 2, 4 e 6.

– Probabilidade de cada face: P(2) = 1/6, P(4) = 1/6, P(6) = 1/6.

– P(par) = P(2 ou 4 ou 6) = P(2) + P(4) + P(6).

– P(par) = 1/6 + 1/6 + 1/6 = 3/6

– P(par) = 0,5 ou P(par) = 50%

Esse exemplo já foi resolvido anteriormente de outra maneira. A comparação das duas formas pode ser ilustrativa na compreensão da propriedade que utilizamos nesta segunda resolução.

7.2.2 Eventos complementares

São eventos complementares aqueles em que a ocorrência de um ou de outro engloba todos os eventos possíveis. Por exemplo: no jogo de moeda, os eventos cara e coroa são complementares. No jogo de dado, teríamos como complementares "face 2" e "não obter face 2".

Quando dois eventos foram complementares, a soma de suas probabilidades será 1.

P(A) + P(não A) = 1

Informalmente, podemos dizer que temos certeza (P = 100%) de que o evento A ocorrerá ou não ocorrerá. Tomando os exemplos anteriores teríamos: a moeda cai cara ou não cai cara; o dado dá a face 2 ou outra face qualquer etc.

É usual em algumas situações chamar de p minúsculo a probabilidade de ocorrência (ou sucesso) e de q a probabilidade de não ocorrer (ou insucesso).

Desse modo, com a soma p + q = 1, temos:

q = 1 - p

Previsão do tempo

Um jornal traz a informação de que a previsão de chuva é de 20%. Qual a probabilidade de não chover?

- Chover e não chover são complementares.
- Chamo a probabilidade de chuva de p, logo, p = 0,2.
- Chamo a probabilidade de não chuva de q.
- q = 1 - p
- q = 1 - 0,2
- q = 0,8

Ou seja, a probabilidade de não chover é de 80%.

7.2.3 Eventos independentes

Chamamos de eventos independentes aqueles cujo acontecimento não interfere na probabilidade de outro acontecer. Quando dois eventos como A e B são independentes, a probabilidade de ambos ocorrerem será dada pelo produto das probabilidades:

P(A e B) = P(A).P(B)

Jogo de dado e moeda

Podemos tomar como exemplo o jogo de uma moeda e de um dado. Em primeiro lugar, verificamos que os eventos são independentes, visto que o resultado do jogo de um não interfere no resultado que se obtém ao jogar o outro. Assim, a probabilidade de ocorrer um resultado específico em uma jogada será dada pelo produto das probabilidades dos dois eventos.

- Qual a probabilidade de o resultado dar cara e cinco?
  - Como já vimos: P(K) = 1/2 e P(5) = 1/6.
  - Probabilidade de faces cara e cinco: P(K e 5) = 1/2 . 1/6.
  - P(K e 5) = 1/12

Para entendermos melhor o resultado: uma vez obtida a face desejada, temos ainda duas possibilidades de resultado da moeda, o que faz a probabilidade do resultado "5 e cara" ser metade da probabilidade de resultado "5". De outro modo, uma vez obtida a face cara, há seis possibilidades de resultado do dado, logo, a probabilidade do resultado específico será um sexto da probabilidade de tirar cara na moeda.

Para qualquer dos dois raciocínios, o resultado será a multiplicação das probabilidades, conforme estabelece a propriedade.

Gênero e compra de computador

O uso da propriedade anterior não se limita a casos analíticos, então, vejamos agora um exemplo com dados de tipo empírico.

Uma loja de informática vende 1 computador a cada 20 clientes que entram na loja. De seus clientes, cerca de 40% são mulheres.

Computadores são vendidos para homens e para mulheres, assim, podemos considerar os eventos independentes para efetuar o cálculo a seguir.

- Qual a probabilidade de um cliente que entre na loja ser uma mulher que compre um computador?
  - Probabilidade de ser mulher: P(M) = 40% ou P(M) = 0,4.
  - Probabilidade de comprar computador: P(c) = 1/20 ou P(c) = 0,05.
  - Como os eventos são independentes, a probabilidade de ambos ocorrerem será o produto das probabilidades de cada evento.
  - P(M e c) = 0,4 * 0,05 = 0,02 ou P(M e c) = 2%.
  - Ou seja, a cada 100 clientes, há 2 mulheres que adquirem um computador.

- Qual a probabilidade de um cliente que entre na loja ser um homem que compre um computador?
  - Eventos "ser homem" e "ser mulher" são complementares.
  - Probabilidade de ser homem: P(M) = 1 - P(H) ou P(M) = 1 - 0,4.
  - P(H) = 0,6
  - Probabilidade de comprar computador: P(c) = 1/20 ou P(c) = 0,05.
  - Como os eventos são independentes, a probabilidade de ambos ocorrerem será o produto das probabilidades de cada evento.
  - P(M e c) = 0,6 * 0,05 = 0,03 ou P(M e c) = 3%
  - Ou seja, a cada 100 clientes, há 3 homens que adquirem um computador.

7.2.4 Eventos não mutuamente excludentes

Quando dois eventos, A e B, podem ocorrer simultaneamente, isso implica em termos interseção não vazia entre os conjuntos A e B. Nesse caso, a probabilidade de ocorrer um ou outro evento será dada por:

P(A ou B) = P(A) + P(B) – P(A e B)

Vimos diversos casos desse tipo na seção 7.4. Essa propriedade traduz formalmente o fato de haver elementos comuns que seriam contados duas vezes.

Convidados da festa

Retomando o exemplo da festa, para calcular a probabilidade de o convidado ser amigo de Pedro ou de Luiz utilizando a propriedade citada, teríamos:

- Total de convidados: $n_{tot} = 70$.
- Probabilidade de ser amigo de Pedro:
  - P(P) = 45/70
  - P(P) = 0,643
- Probabilidade de ser amigo de Luiz:
  - P(P) = 35/70
  - P(P) = 0,5
- Probabilidade de ser amigo de ambos:
  - P(P) = 15/70
  - P(P) = 0,214
- Probabilidade de ser amigo de Pedro ou de Luiz:
  - P = 0,643 + 0,5 - 0,214
  - P = 0,93

Para ilustrar o significado da probabilidade obtida, podemos utilizar os dados do exemplo em que vimos haver, das 70 pessoas presentes na festa, apenas 5 que não são amigas de nenhum deles, ou seja, há 65 amigos de um dos dois, logo:

- Probabilidade de ser amigo de Pedro ou de Luiz:
  - P = 65/70
  - P = 0,93

Obviamente, como tínhamos acesso aos números de convidados, esse caminho é mais fácil de utilizar, no entanto, o objetivo aqui era o de ilustrar a validade da propriedade. Além disso, quando temos somente as probabilidades, sua utilização é necessária.

7.3 Distribuições de probabilidade

Para dar continuidade a nosso estudo de probabilidades, vamos agora discutir como se constroem alguns modelos estatísticos que trazem como informação as probabilidades de todos os eventos possíveis para certo universo estudado, as chamadas distribuições de probabilidade.

Nesta disciplina estudaremos duas das principais distribuições: a binomial e a normal.

Como vimos no cálculo de probabilidades, muitas vezes é preciso analisar a situação de modo a podermos contar ou número de eventos. Veremos, a seguir, como realizar tais contagens em uma série de situações, para que mais adiante possamos compreender a primeira distribuição de probabilidades que abordaremos: a distribuição binomial.

A outra distribuição a ser estudada, chamada de normal, retomará os conceitos de média e desvio-padrão vistos anteriormente e os utilizará para estimar probabilidades de faixas de valores. Para facilitar a compreensão, discutiremos antes dela alguns conceitos importantes a partir da distribuição uniforme.

7.4 Análise combinatória

A análise combinatória é utilizada para realizar a contagem do número de eventos em uma situação dada. Seu estudo nos permitirá, ao final, ter o ferramental matemático necessário para compreender a distribuição binomial.

7.4.1 Princípio da multiplicação

Nesta seção vamos discutir algumas situações que ilustram o fato de utilizarmos multiplicações para a contagem dos eventos.

Exemplo 1: Combinando roupas

Comecemos a estudar a análise combinatória a partir de um exemplo simples: determinar quantos conjuntos diferentes se podem obter com 2 calças e 5 blusas.

Temos cinco possibilidades de escolha de calças e, para cada calça, há duas possibilidades de escolha de blusa. O quadro a seguir sintetiza essa análise:

$$\underbrace{\boxed{5}}_{\text{calças}} * \underbrace{\boxed{2}}_{\text{blusas}} = \underbrace{\boxed{10}}_{\text{conjuntos}}$$

Assim sendo, há 10 possibilidades de combinarmos roupas utilizando essas peças.

Exemplo 2: Arrumação de mesa de festa

Suponha agora que um bufê tem disponíveis, para a arrumação das mesas, três tipos de pratos, um faqueiro prateado e outro dourado e quatro modelos de toalhas. De quantas maneiras diferentes se podem arrumar as mesas usando esses artigos?

Para responder a essa pergunta, novamente recorremos ao princípio da multiplicação:

5	*	2	*	4	=	40
pratos		talheres		toalhas		conjuntos

Há, portanto, 40 possibilidades de arrumação de mesa.

Para ilustrarmos a utilização desse tipo de resultado no cálculo de probabilidades, verificamos que a probabilidade de um cliente vir a escolher certa combinação específica será dada por:

- P = 1/40, logo P = 0,0025 ou P = 2,5%.

Note que o resultado anterior supõe que todas as arrumações fiquem igualmente harmônicas, de modo tal que a escolha de todas elas seja igualmente provável.

Exemplo 3: Prova de múltipla escolha

Outro exemplo de contagem dessa forma é calcular quantas são as possibilidades de respostas em uma prova de múltipla escolha.

Suponhamos uma prova de 4 questões, com 5 alternativas cada uma.

Utilizando o princípio da multiplicação, teremos:

5	*	5	*	5	*	5	=	625
questão 1		*questão 2*		*questão 3*		*questão 4*		*total*

Reescrevendo a conta acima de maneira mais sintética, e chamando o número total de *resultados possíveis de N*, teríamos:

$N = 5^4$, logo N = 625.

A partir disso, podemos calcular qual a probabilidade de um aluno, que escolha as respostas de maneira aleatória, acertar todas as questões:

- Existe apenas um conjunto de respostas corretas.
- Existem 625 possibilidades de respostas.

- $P = 1/625$, logo $P = 0,0016$ ou $P = 0,16\%$.

Suponhamos agora que a prova tenha 10 questões. Qual a probabilidade de tirar 10 na prova sem saber nada da matéria?

- O número total de resultados será: $N = 5^{10}$.
- O número total de resultados será: $N = 9765625$.
- $P = 1/9765625$, logo $P = 0,000000102$ ou $P = 0,0000102\%$.

Ou seja, se houver 10 questões, a probabilidade será de aproximadamente 1 em 10 milhões.

7.4.2 Permutações

Passaremos agora a utilizar a técnica de contagem vista anteriormente para analisar casos de permutações, ou seja, em que há um conjunto de elementos cuja ordem será alterada trocando-se suas posições.

7.4.2.1 Permutas sem repetição

Como exemplo desse caso, pergunta-se: de quantas maneiras posso arrumar 4 livros em uma prateleira?

Para responder a essa pergunta, uma maneira de pensar é a seguinte:

- Há 4 livros possíveis para a primeira posição.
- Para a segunda posição, como já foi colocado um livro, tenho somente 3 escolhas.
- Para a terceira posição, haverá 2 escolhas possíveis.
- A quarta posição será ocupada pelo livro restante.

Utilizando o mesmo esquema da seção anterior, teremos:

4	*	3	*	2	*	1	=	24
posição 1		*posição 2*		*posição 3*		*posição 4*		*total*

Utilizando o mesmo tipo de raciocínio, responda: de quantas maneiras posso arrumar 8 livros em uma prateleira?

A resposta a essa pergunta pode ser obtida seguindo o mesmo procedimento descrito anteriormente, dessa forma, teremos:

$N = 8 * 7 * 6 * 5 * 4 * 3 * 2 * 1$

$N = 40320$

Poderíamos utilizar esse procedimento para qualquer número de livros, no entanto, quando o número de elementos cresce, torna-se cada vez menos prático indicar a conta de maneira explícita. Para evitar a necessidade de se escrever contas muito grandes, utiliza-se a notação fatorial.

Na notação fatorial, um ponto de exclamação após o número indica a multiplicação sucessiva do número pelo número anterior até que se chegue ao número 1. Ou seja:

- 4! = 4 * 3 * 2 * 1
- 8! = 8 * 7 * 6 * 5 * 4 * 3 * 2 * 1

Utilizando essa notação, então teremos para permutas, num caso geral, o número de resultados possíveis para sequências que envolvam n elementos será dado por:

$P_n = n!$

Vale ressaltar o caso especial em que n = 0:

0! = 1

Vejamos um exemplo que ilustra esse resultado: se tivermos n = 0, ou seja, nenhum livro para colocar na prateleira, haverá apenas uma possibilidade: deixar a prateleira vazia.

7.4.2.2 Permutas com repetição

Nos exemplos anteriores, supusemos sempre que os livros em questão eram diferentes. Mas, se houver livros iguais, haverá sequências repetidas entre as que contamos. Assim sendo, nos casos em que há elementos repetidos, é preciso dividir o número de permutações encontradas pelo número de vezes em que cada uma aparece.

Tomemos como exemplo o caso em que temos 5 livros, sendo que 2 deles são iguais.

- Primeiro, fazemos as contas sem nos preocuparmos com as repetições:
  - P_5 = 5! = 5 * 4 * 3 * 2 * 1 = 120
  - P_5 = 5 * 4* 3 * 2 * 1
  - P_5 = 120
- Para analisar as repetições, escolhemos uma sequência determinada qualquer para analisar quantas iguais a ela haverá:
  - Como são 2 iguais, tenho 2 posições com livros iguais.
  - Para a primeira posição, tenho duas escolhas de livro.

- O segundo livro vai no lugar restante.
- Ou seja, serão 2 repetições.

- Como cada resultado aparece 2 vezes, tenho, em realidade, metade das sequências inicialmente supostas.

  - Assim:

  $N = P_5/2$

  $N = 120/2$

  $N = 60$

Façamos agora o mesmo raciocínio para o caso de haver 3 livros iguais:

- Novamente começamos sem nos preocupar com as repetições:

  - $P_5 = 120$

- Para analisar as repetições, escolhemos uma sequência determinada qualquer para analisar quantas iguais a ela haverá:

  - Como são 3 iguais, tenho 3 posições com livros iguais.
  - Para a primeira posição, tenho 3 escolhas de livro.
  - Para a segunda posição, tenho 2 escolhas de livro.
  - O terceiro livro igual vai no lugar restante.
  - Ou seja, serão 3! repetições.

- Como cada resultado aparece 3! vezes, tenho, em realidade, um sexto das sequências inicialmente supostas.

  - Assim:

  $N = P_5/3!$

  $N = 120/6$

  $N = 20$

Finalmente, façamos a mesma análise para o caso de haver 3 livros de um tipo e 2 livros de outro.

- Sem nos preocuparmos com as repetições:

  – $P_5 = 120$

- Temos agora duas classes de repetições, uma referente ao livro que aparece 2 vezes, outra referente ao livro que aparece 3 vezes.

  – Quando há 2 iguais, o número de repetições é 2!

  – Quando há 3 iguais, o número de repetições é 3!

  – O número total de repetições será o produto de 2! por 3!

- O resultado final será, portanto, o número sem repetições dividido por 2! e também por 3!.

  – Assim:

  $N = P_5/(2!3!)$

  $N = 120/(2*6)$

  $N = 120/(12)$

  $N = 10$

Generalização e formalização

Para efetuar a contagem do número de sequências possível na permuta de n elementos temos:

$P_n = n!$

Lembrando que $0! = 1$

Se houver elementos repetidos, é preciso considerar esse fato dividindo o resultado inicial pelo número de repetições, como segue:

$$P_n^{a,b,c...} = \frac{n!}{a!b!c!...}$$

Onde a, b e c são o número de vezes em que os diferentes elementos A, B e C se repetem no conjunto.

Exemplos adicionais: anagramas

Anagramas são rearranjos de letras de uma dada palavra. O cálculo do número de anagramas que se pode formar com as letras de uma dada palavra são um exemplo clássico de utilização do cálculo de permutações. Vejamos alguns:

- Calcular os anagramas da palavra "lápis".

  – Temos 5 letras, sem repetição.

  – $N = P_5$

  – $N = 5!$

  – $N = 120$

- Calcular os anagramas da palavra "papel".

  – Temos 5 letras, com a letra **p** aparecendo 2 vezes.

  – $N = P_5^2$

  – $N = \frac{5!}{2!}$

  – $N = \frac{120}{2}$

  – $N = 60$

- Calcular os anagramas da palavra "telefone".

– Temos 8 letras, com a letra **e** aparecendo 3 vezes.

  – $N = P_8^3$

  – $N = \frac{8!}{3!}$

  – $N = \frac{40320}{6}$

  – $N = 6720$

- Calcular os anagramas da palavra "comodidade".

– Temos 10 letras, com a letra **d** aparecendo 3 vezes e a letra **o** duas vezes.

  – $N = P_{10}^{3,2}$

  – $N = \frac{10!}{3!2!}$

– $N = \frac{3628800}{6 * 2}$

– $N = 302400$

7.4.3 Arranjos

Vamos agora analisar situações similares às permutas, mas com a seguinte característica adicional: há mais elementos que posições a serem ocupadas.

Por exemplo: quantos resultados possíveis há num campeonato em que participem 8 atletas, considerando as 3 primeiras posições?

Para responder a isso, procedemos de maneira similar aos exemplos referentes à permuta.

Uma maneira de pensar o problema é a seguinte:

- Há 8 atletas possíveis para a primeira posição.
- Para a segunda posição, como já foi atribuída a medalha de ouro, tenho somente 7 possibilidades.
- Definidas as medalhas de ouro e prata, restarão 6 atletas possíveis para a medalha de bronze.

Utilizando o mesmo esquema da seção anterior, teremos:

$$\underset{ouro}{\boxed{8}} \ * \ \underset{prata}{\boxed{7}} \ * \ \underset{bronze}{\boxed{6}} \ = \ \underset{total}{\boxed{336}}$$

Ou seja, embora a compreensão disso seja semelhante à do caso das permutações, dado que nesse caso utilizamos uma lógica similar, é preciso buscarmos uma maneira de representar esse raciocínio para casos gerais.

Imagine como seria escrever uma situação em que houvesse 200 elementos para ocupar 150 posições. Ainda que o raciocínio seja o mesmo, escrever o produto de 150 números não é nada prático. Vamos, então, buscar uma maneira de indicar a conta a ser feita de modo sintético.

- Como temos um produto de números sequenciais, é natural pensarmos que a formalização utilizará a notação fatorial.
- Como esse produto vai ser truncado em algum ponto, é preciso indicar esse truncamento, ou seja, indicar até que número vai a multiplicação.

A maneira de escrever de forma apropriada a descrição dada acima, para um caso geral em que haja n elementos para **p** posições, é a seguinte:

$$A_{n,p} = \frac{n!}{(n-p)!}$$

Retomemos o exemplo do campeonato para compreendermos melhor a utilização e o significado dessa fórmula.

- Em primeiro lugar, identificaremos os valores de n e de p:
  - n indica o total de elementos participantes, logo, n = 8.
  - p indica o total de posições a ocupar, logo, n = 3.
- De posse dos valores de n e p, podemos utilizar a fórmula:
  - Substituímos os valores de n e de p, como segue:

$$A_{8,3} = \frac{8!}{(8-3)!}$$

  - Agora, efetuamos a conta indicada no denominador:

$$A_{8,3} = \frac{8!}{5!}$$

  - Em seguida, expandimos os fatoriais:

$$A_{8,3} = \frac{8*7*6*5*4*3*2*1}{5*4*3*2*1}$$

  - Então, dividimos os termos que aparecem no numerador e também no denominador (ao que chamamos comumente "cortar"):

$$A_{8,3} = \frac{8*7*6*\cancel{54321}}{\cancel{54321}}$$

  - Com isso, chegamos à expressão que havíamos deduzido diretamente no início desta seção:

$$A_{8,3} = 8*7*6$$
$$A_{8,3} = 336$$

Ou seja, a fórmula geral dada exprime exatamente o mesmo raciocínio que foi exposto anteriormente, mas utilizando uma notação formal e genérica que permite a descrição de qualquer arranjo.

7.4.4 Combinações

Neste item, vamos analisar as combinações. Esses casos referem-se a situações similares aos arranjos em que a ordem seja é indiferente.

Tomando o exemplo da seção anterior, imaginemos que a ordem de classificação não importasse, que em lugar de 3 medalhas fossem 3 vagas para as olimpíadas. Desse modo, importariam somente quais seriam os classificados e teríamos situação equivalente quando os mesmos 3 atletas fossem selecionados para formar a equipe. Nossa contagem consideraria isso dividindo a contagem referente ao arranjo pelo número de possibilidades de haver um mesmo trio classificado, ou seja, temos que dividir por 3! o resultado do arranjo, como veremos a seguir.

Pergunta-se, portanto, de quantas maneiras se pode formar uma equipe de 3 atletas escolhendo 3 entre 8?

Para responder a isso, começamos novamente os cálculos de maneira similar aos exemplos referentes à permuta.

- Há 8 atletas possíveis para a primeira posição, 7 para a segunda e 6 para a terceira.

8	*	7	*	6	=	336
vaga 1		*vaga 2*		*vaga 3*		*total*

- Como todas as vagas são equivalentes, a equipe ABC é a mesma que BCA, por exemplo.
  - O número de repetições no caso em que há 3 posições será dado por 3 * 2 * 1, ou seja, 3!
  - Desse modo, nossa contagem fica sendo: $N = \frac{8 * 7 * 6}{3!}$
  - Efetuando as multiplicações, temos $N = \frac{336}{6}$
  - Logo, N = 56.

Novamente, para que seja viável descrever qualquer situação similar, é preciso buscar uma maneira de representar esse raciocínio para casos gerais, a exemplo do que fizemos para os arranjos.

Como a diferença entre a contagem dos arranjos e a das combinações se dá pela divisão do número de repetições, então a maneira apropriada de escrever a contagem descrita anteriormente para um caso geral em que haja n elementos para p posições é a seguinte:

$$C_{n,p} = \frac{n!}{p!(n-p)!}$$

Retomemos o exemplo para compreender melhor a utilização dessa fórmula.

- Em primeiro lugar, identificaremos os valores de n e de p:
  - n indica o total de elementos participantes, logo, n = 8.
  - p indica o total de posições a ocupar, logo, p = 3.

- De posse dos valores de n e p, podemos utilizar a fórmula:

  – Substituímos os valores de n e de p, como segue:

$$C_{8,3} = \frac{8!}{3!(8-3)!}$$

  – Agora, efetuamos a conta indicada no denominador e teremos:

$$C_{8,3} = \frac{8!}{3!5!}$$

  – Em seguida, expandimos os fatoriais:

$$A_{8,3} = \frac{8*7*6*5*4*3*2*1}{3*2*1*5*4*3*2*1}$$

  – Então, dividimos os termos que aparecem no numerador e também no denominador:

$$A_{8,3} = \frac{8*7*6*\cancel{5}*\cancel{4}*\cancel{3}*\cancel{2}*\cancel{1}}{3*2*1*\cancel{5}*\cancel{4}*\cancel{3}*\cancel{2}*\cancel{1}}$$

  – Com isso, chegamos à expressão que havíamos deduzido diretamente no início desta seção:

$$C_{8,3} = \frac{8*7*6}{3*2*1}$$
$$C_{8,3} = 56$$

7.5 Distribuições de probabilidade

As distribuições de probabilidade são modelos estatísticos que descrevem as probabilidades dos diversos resultados de um determinado universo ou espaço amostral.

Assim como no estudo de análise combinatória, há situações de interesse que possuem características tais que permitem um tratamento comum. Na contagem, os casos vistos foram as permutas, os arranjos e as combinações.

Nas distribuições de probabilidade veremos também dois casos de maior interesse: a distribuição binomial e a distribuição normal.

Observação

Há diversos tipos de distribuição de probabilidades, mas não buscamos um estudo exaustivo do tema, vamos enfocar somente os dois de maior interesse.

7.5.1 Variáveis aleatórias

Uma variável aleatória é aquela cujo valor depende de fatores imponderáveis do acaso.

São variáveis aleatórias o resultado de um sorteio, a quantidade de pessoas que passará numa rua num determinado horário, o horário em que a chuva começa etc.

7.5.2 Tipos de distribuição

Podemos classificar a distribuição de probabilidades de acordo com o caráter de sua variável.

Distribuições discretas

Quando a variável aleatória em estudo é discreta, a distribuição de probabilidades correspondente também será discreta.

A distribuição de probabilidades, nesse caso, será dada pela probabilidade de cada um dos valores possíveis específicos.

Existem diversas distribuições discretas, dadas pela natureza da situação estudada. Neste curso estudaremos somente a distribuição binomial.

Distribuições contínuas

Quando a variável aleatória em estudo é contínua, a distribuição de probabilidades correspondente também será contínua.

Quando a variável é contínua, não tratamos mais de probabilidades referentes a valores individuais, mas sim referentes a intervalos de valores. A distribuição de probabilidades nesses casos nos dará a probabilidade de encontrar o valor da variável dentro de certo intervalo.

Também existem diversas distribuições contínuas, dadas pela natureza da situação estudada, mas, em virtude da natureza desta disciplina, abordaremos apenas a distribuição normal.

Observação

O tipo de distribuição dependerá, entre outros fatores, do tipo de grandeza que estamos estudando.

7.6 Distribuição binomial

A distribuição binomial se aplica a situações que possuem as seguintes características:

- A variável aleatória é discreta.
- n eventos são independentes.
- Para cada evento, dois resultados possíveis e complementares (sucesso e insucesso).
  - Probabilidade de sucesso (de ocorrer o evento escolhido): p.
  - Probabilidade de insucesso (de não ocorrer o evento escolhido): q.
  - $q = 1 - p$

Quando esses requisitos são atendidos, a distribuição binomial nos dá a maneira de calcular a probabilidade de que o evento escolhido ocorra um número p de vezes, sendo que essa quantidade poderá variar entre 0 e n.

Buscando construir a distribuição, vamos utilizar um exemplo de fácil visualização. Depois de formalizado o cálculo, resolveremos outros exemplos.

7.6.1 Jogo de n moedas

Para construir a distribuição binomial, tomemos como exemplo o jogo de n moedas. Note-se, inicialmente, que esse exemplo atende aos requisitos listados anteriormente.

No jogo de uma única moeda, o resultado da jogada é um imponderável. Podemos escolher como variável aleatória o número de faces cara obtidas, ao que chamaremos k. Em n jogadas ou eventos, teremos certo número k de moedas que terminem com a face cara para cima, variando entre 0 e n. Assim, a distribuição de probabilidades consistirá em analisar qual a probabilidade de obtermos cada um dos resultados k.

O resultado "face cara" será considerado um sucesso, e sua probabilidade será $p = 0,5$. O resultado "face coroa" significa a não ocorrência de "face cara", logo, o chamaremos de insucesso, com probabilidade q. Como os eventos são complementares, temos $q = 1 - p$ e, portanto $q = 0,5$.

Utilizando essas informações buscaremos a expressão geral a partir das expressões referentes a $n = 1$, $n = 2$ e $n = 3$.

Jogo de 1 moeda ($n = 1$)

Como estamos considerando o jogo de somente uma moeda, os valores possíveis dessa variável serão $k = 0$ e $k = 1$. Formalmente, podemos construir a distribuição de probabilidades para a obtenção de face cara no jogo de uma moeda da seguinte forma:

- Probabilidade de nenhuma face cara:

  - $k = 0$

  - $P(0) = p$, logo, $P(0) = 0{,}5$.

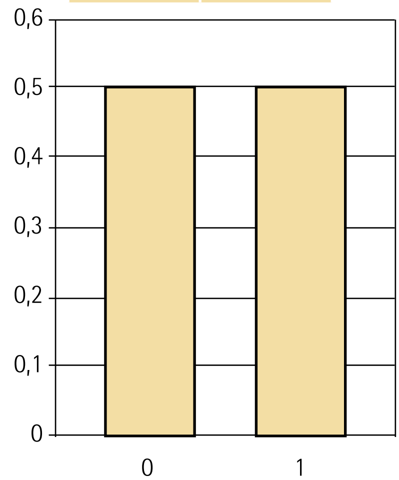
- Probabilidade de uma face cara:

  - $k = 1$

  - $P(1) = q$, logo, $P(1) = 0{,}5$.

Em resumo:

k	P(k)
0	0,5
1	0,5



Jogo de 2 moedas (n = 2)

Como temos 2 moedas, os valores da variável escolhida serão $k = 0$, $k = 1$ e $k = 2$. Lembrando que os eventos são independentes, visto que o fato de uma moeda cair com certa face para cima não interferirá na face que ficará para cima da outra moeda, então a probabilidade de um resultado específico para o conjunto será dada pelo produto das probabilidades de cada resultado específico.

Ou seja, se escolho como resultado que a primeira moeda terá "face cara" e a segunda moeda "face coroa", a probabilidade desse resultado específico do conjunto será dada pelo produto da probabilidade de a primeira ser cara (p) e a probabilidade de a segunda ser coroa (q), o que se escreve como:

P = p.(1 - q)

P = 0,5*(1 - 0,5)

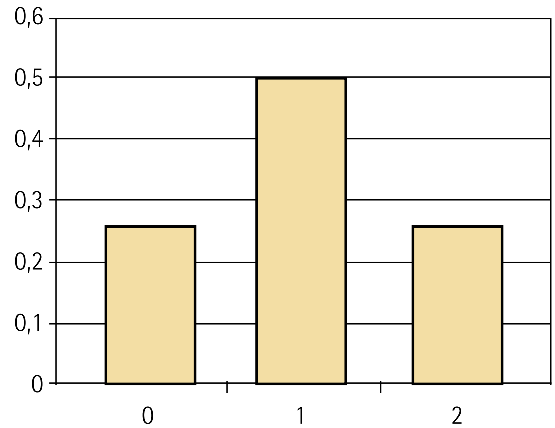
P = 0,25

Passamos agora a construir a distribuição nesse caso.

- Probabilidade de nenhuma face cara (k = 0):
  - 2 insucessos: P = q.q = 0,5*0,5 = 0,25
  - 1 maneira de obter 2 insucessos: CC.
  - P(0) = 1*0,25
  - P(0) = 0,25
- Probabilidade de uma face cara (k = 1):
  - 1 sucesso e 1 insucesso: P = p.q = 0,5*0,5 = 0,25.
  - 2 maneiras de obter 1 sucesso: KC ou CK.
  - P(1) = 2*0,25
  - P(1) = 0,5
- Probabilidade de duas faces cara (k = 2):
  - 2 sucessos: P = p.p = 0,5*0,5 = 0,25.
  - 1 maneira de obter 2 sucessos: KK.
  - P(0) = 1*0,25
  - P(0) = 0,25

Em resumo:

k	P(k)
0	0,25
1	0,5
2	0,25



Jogo de 3 moedas (n = 3)

Como temos 3 moedas, os valores da variável escolhida serão agora k = 0, k = 1, k = 2 e k = 3.

A distribuição de probabilidades nesse caso será:

- Probabilidade de nenhuma face cara (k = 0):
  - 3 insucessos: P = q.q.q = q^3 = 0,5*0,5*0,5 = 0,125.
  - 1 maneira de obter 3 insucessos: CCC.
  - P(0) = 1*0,125
  - P(0) = 0,125
- Probabilidade de uma face cara (k = 1):
  - 1 sucesso e 2 insucessos: P = p.q.q = p.q^2= 0,5*0,5*0,5 = 0,125.
  - 3 maneiras de obter 1 sucesso: KCC ou CKC ou CCK.
  - P(1) = 3*0,125
  - P(1) = 0,375
- Probabilidade de duas faces cara (k = 2):
  - 2 sucessos: P = p.p.q = p^2.q = 0,5*0,5*0,5 = 0,125.
  - 3 maneiras de obter 2 sucessos: KKC ou KCK ou CKK.
  - P(0) = 3*0,125
  - P(0) = 0,375

- Probabilidade de três faces cara ($k = 3$)

– 3 sucessos: $P = p.p.p = p^3 = 0,5 * 0,5 * 0,5 = 0,125$

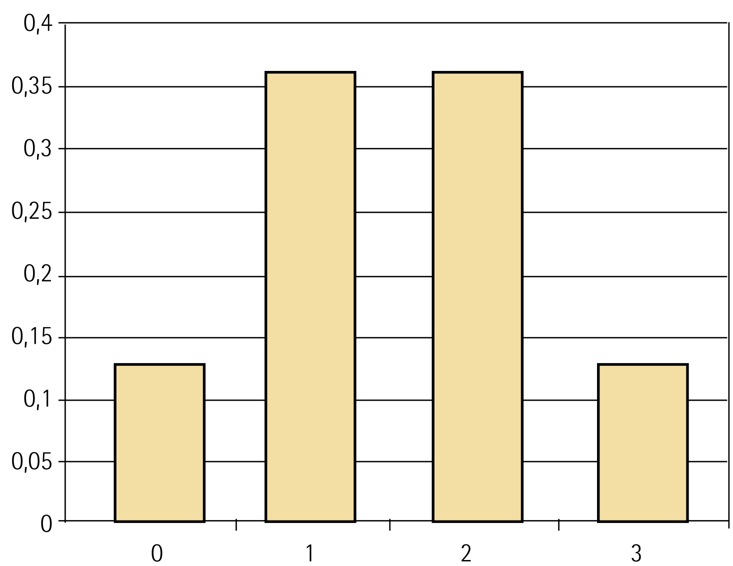
– 1 maneira de obter 3 sucessos: KKK.

– $P(0) = 1 * 0,125$

– $P(0) = 0,125$

Em resumo:

k	P(k)
0	0,1250
1	0,3750
2	0,3750
3	0,1250



Jogo de n moedas

Assim como feito anteriormente, para obtermos a distribuição de probabilidades se tivermos n moedas, temos que analisar a probabilidade de termos resultados que variam entre $k = 0$ e $k = n$.

Para isso, primeiro temos que avaliar a probabilidade de uma jogada específica.

- Se temos k sucessos, o número de insucessos será $n - k$, ou seja, o restante das moedas.

- A probabilidade será dada pelo produto das probabilidades p (k vezes) e q ($n - k$ vezes) referentes a cada moeda. Ou seja, os expoentes das probabilidades serão o número de vezes que o evento ocorre. Ou seja, a probabilidade de sucessos, p, terá expoente k. A probabilidade de insucessos, q, terá expoente correspondente ao restante das vezes, $n - k$:

$$P = p^k * q^{n-k}$$

Em seguida, é preciso avaliar quantas vezes cada resultado específico aparecerá:

- São n moedas para serem colocadas em k lugares.
- A ordem em que as moedas com o resultado escolhido aparece não é relevante.
- Temos, portanto, um caso de combinação, e o número de maneiras que podemos obter um certo valor k será dado por:

$$C_{n,k} = \frac{n!}{k!(n-k!)}$$

A probabilidade referente a certo valor de k será p, produto da probabilidade de uma vez pelo número de maneiras que esse resultado aparece.

$$P(k) = \frac{n!}{k!(n-k!)} p^k q^{n-k}$$

Utilizando essa fórmula geral para um caso já calculado, o jogo de 3 moedas com 2 resultados cara, teremos:

- Identificação dos termos da expressão:
  - Número de moedas: n = 3.
  - Número de sucessos: k = 2.
  - Probabilidade de sucessos: p = 0,5.
  - Probabilidade de insucessos: q = 1 - 0,5, q = 0,5.
- Substituição dos termos na expressão:
  - $P(2) = \frac{3!}{2!(3-2!)} * 0{,}5^2 * 0{,}5^{(3-2)}$
- Realizar operações:
  - $P(2) = \frac{3!}{2!(1!)} 0{,}5^2 * 0{,}5^1$
  - $P(2) = \frac{3*2*1}{2*1*1} 0{,}25 * 0{,}5$
  - $P(2) = \frac{6}{2} * 0{,}125$

– P(2) = 3*0,125

– P(2) = 0,375

7.6.2 Caso geral

É comum representar a fórmula do cálculo de combinações por meio da representação do binômio de Newton, de onde se origina a denominação da distribuição. O binômio é dado por:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Utilizando essa notação e escrevendo explicitamente q como 1 - p, a expressão encontrada passa a ser escrita assim:

$$P(k) = \binom{n}{k} p^n (1-p)^{n-k}$$

Vale notar que, em um caso geral, ao contrário do que tínhamos no exemplo da moeda, as probabilidades de sucesso e insucesso podem ser diferentes.

Por exemplo, se tivermos n = 5, k = 2 e p = 0,1, o cálculo da probabilidade será realizado da seguinte forma:

- $P(2) = \binom{5}{2} 0{,}1^5 (1-0{,}1)^{5-2}$
- $P(2) = \frac{5!}{2! * (5-2)!} 0{,}1^5 * 0{,}9^3$
- $P(2) = \frac{120}{2 * 6} * 0{,}1^5 * 0{,}9^3$
- $P(2) = 10 * 0{,}00001 * 0{,}73$
- P(2)=0,0729

Observação

Na prática, atualmente, os valores do binômio são geralmente calculados com a utilização de planilhas eletrônicas.

Média e desvio-padrão

A média de valores, quando temos uma distribuição binomial, é dada por:

$\overline{x} = n.p$

O desvio-padrão é dado por:

$\sigma = n.p.q$

Exemplo

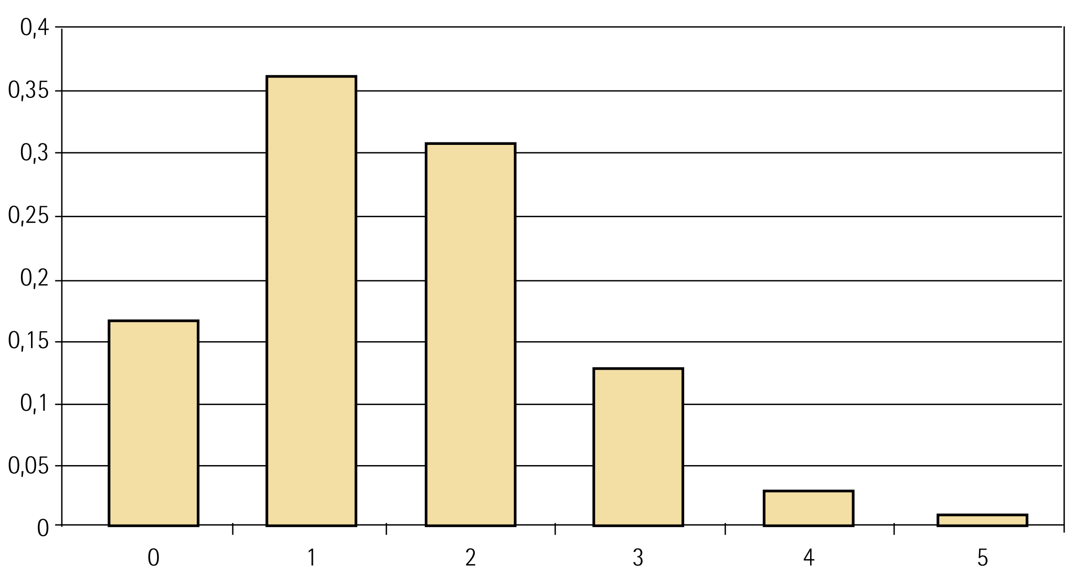
Vamos analisar um caso em que temos n = 5 e p = 0,3.

Em primeiro lugar, podemos calcular a média e o desvio-padrão:

$\overline{x} = 5 * 0,3 = 1,5$ e $\sigma = 5 * 0,3 * 0,7 = 1,05$

Agora, vejamos a distribuição de probabilidades. O procedimento para o cálculo já foi apresentado no exemplo anterior de maneira detalhada, assim, apresentaremos somente os resultados finais para cada valor de k.

k	P(k)
0	0,1681
1	0,3602
2	0,3087
3	0,1323
4	0,0284
5	0,0024



Note que, ao contrário do exemplo da moeda, agora o gráfico não é mais simétrico. Isso acontece porque a probabilidade de o evento ocorrer é menor que a de não ocorrer, assim, as probabilidades relacionadas a um maior número de ocorrências passa a ser menor.

7.6.3 Aplicação

Vamos agora construir e analisar alguns exemplos de distribuições binomiais de probabilidade para algumas escolhas de n e p.

O procedimento para o cálculo já foi apresentado no exemplo anterior de maneira detalhada, assim, os exemplos a seguir trarão somente os resultados finais para cada valor de k.

Defeitos em produtos

Estima-se que cerca de 5% das cadeiras produzidas por um fabricante apresentem algum tipo de defeito. Em um lote de 10 cadeiras, qual a probabilidade de não haver nenhuma defeituosa?

- Identificação dos termos da expressão:
  - Número de eventos (cadeiras no lote): n = 10.
  - Número de sucessos (cadeiras com defeito): k = 0.
  - Probabilidade de sucessos (proporção de defeitos): p = 0,05.
  - Probabilidade de insucessos: q = 1 - 0,05, q = 0,95.
- Substituição dos termos na expressão:
  - $P(0) = \frac{10!}{0!(10-0)!} * 0{,}05^0 * 0{,}95^{(10-0)}$
- Realizar operações:
  - Lembrando que 0! = 1 e $0{,}5^0 = 1$: $P(0) = \frac{10!}{1 * 10!} * 1 * 0{,}6$
  - $P(0) = \frac{10!}{10!} * 0{,}6$
  - P(0) = 0,6

Ou seja, com uma proporção de 5% de peças defeituosas, um lote de 10 cadeiras tem a probabilidade de não ter nenhuma com defeito de 60%.

Outra pergunta relevante e bastante fácil de responder, utilizando-se as propriedades da distribuição binomial, é o número médio de cadeiras com defeito num lote de 10 peças.

É óbvio que não se pode ter meia cadeira com defeito. O significado do número anterior fica mais claro se pensarmos que, em média, a cada 2 lotes deve haver uma cadeira com defeito.

Atrasos de pagamento

Uma prestadora de serviços verificou que cerca de 10% dos 20 clientes de uma certa modalidade de atendimento costuma atrasar o pagamento de suas prestações.

Um primeiro número de interesse é saber qual a média de clientes em atraso a cada mês.

$\overline{x} = n.p = 20 * 0,1$

$\overline{x} = 2$

E qual a probabilidade de não termos atrasos acima da média?

Para responder a isso, precisamos inicialmente entender que se o número de sucessos (atrasos) deve ser de, no máximo, 2, esse número pode ser 0 ou 1 ou 2, logo, tenho que fazer as contas para k = 0, k = 1 e k = 2 e somar as probabilidades encontradas.

Para efetuar os cálculos, seguimos o procedimento já ilustrado anteriormente:

- Identificação dos termos da expressão:
  - Número de eventos (clientes): n = 20.
  - Probabilidade de sucesso: p = 0,1.
  - Probabilidade de insucessos: q = 1 - 0,1, q = 0,9.
- Cálculo de P(0), P(1) e P(2) substituindo os valores correspondentes na fórmula, conforme já discutido.

Novamente, serão apresentados os resultados finais para cada valor de k, com aproximação de 4 casas decimais:

k	P(k)
0	0,1216
1	0,2702
2	0,2852

P(k<3) = 0,1216 + 0,2702 + 0,2852

P(k<3) = 0,677

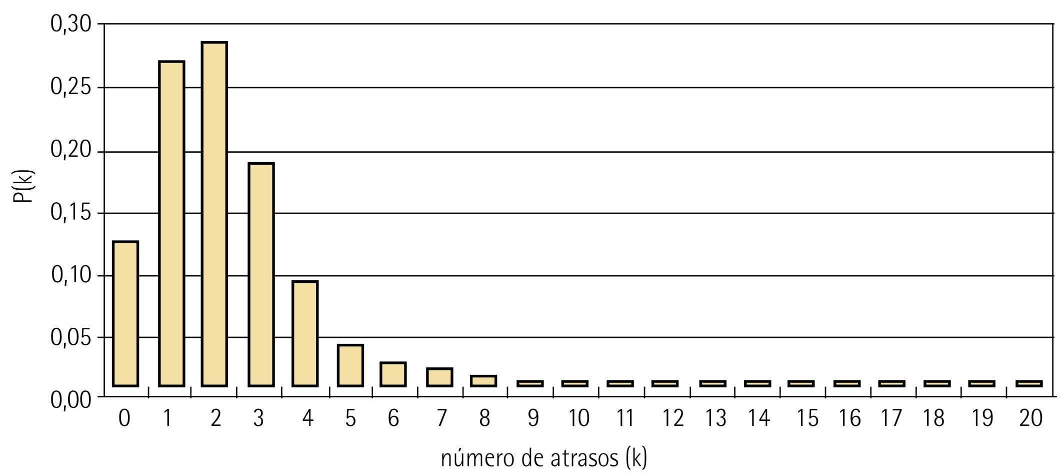
Ou seja, há uma probabilidade de quase 70% de o número de atrasos não ser maior que a média.

Com relação a esse exemplo, vale tecer um comentário a respeito da utilização desses dados.

Uma probabilidade de 70% de os atrasos não serem maiores que a média significa também que há uma probabilidade de 30% de haver atrasos acima da média. Se um gestor conta com esses pagamentos para efetuar pagamentos e avalia somente o número médio de atrasos, contando com 18 pagamentos, haverá uma probabilidade de aproximadamente 30% de não haver o dinheiro disponível dentro desse planejamento.

Não poder contar com a entrada de dinheiro programada implica, muitas vezes, em pagamentos de juros e multas, o que pode comprometer o orçamento da empresa. Como, então, trabalhar com segurança?

No gráfico a seguir, apresentamos a distribuição de probabilidades para esse exemplo:



O gráfico nos mostra que a probabilidade de haver mais de 7 atrasos é praticamente inexistente, sendo que podemos trabalhar com alto grau de segurança se contarmos com um número máximo de 5 atrasos de pagamento.

Uma análise desse tipo pode evitar, portanto, grandes problemas de fluxo de caixa.

7.7 Distribuição uniforme de probabilidades

Uma distribuição de probabilidades uniforme é aquela em que a variável aleatória é contínua e todos os valores são igualmente prováveis. Para facilitar a compreensão, vamos introduzir alguns aspectos das distribuições contínuas de probabilidade por meio de um exemplo bastante intuitivo.

Imagine que, para realizar um sorteio em uma festa, uma semente pequena e dura seja acrescentada a uma massa de bolo. O ganhador do sorteio seria aquele que a encontrasse em sua fatia.

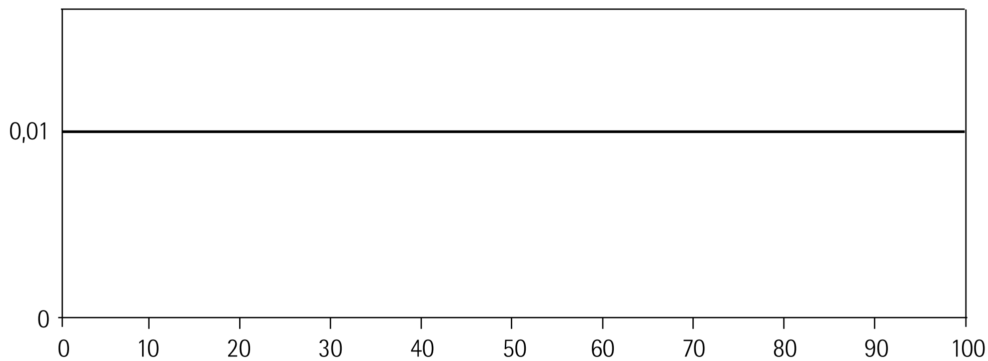
Como não existe motivo algum para que haja uma posição preferencial da semente dentro da massa, podemos assumir que todos os pontos do bolo têm igual probabilidade de conter a dita semente e, portanto, temos um caso de distribuição uniforme de probabilidades.

Supondo que o bolo seja retangular e tenha 1 metro de comprimento, a cada centímetro de bolo teríamos uma probabilidade de 1% de encontrar a semente.

Formalmente, diríamos que:

$f(x) = 0,01$

A distribuição uniforme de probabilidades, em razão da posição, está descrita no gráfico a seguir:



Teríamos, então a situação em que a probabilidade de encontrar a semente seria tanto maior quanto mais larga fosse a fatia de bolo.

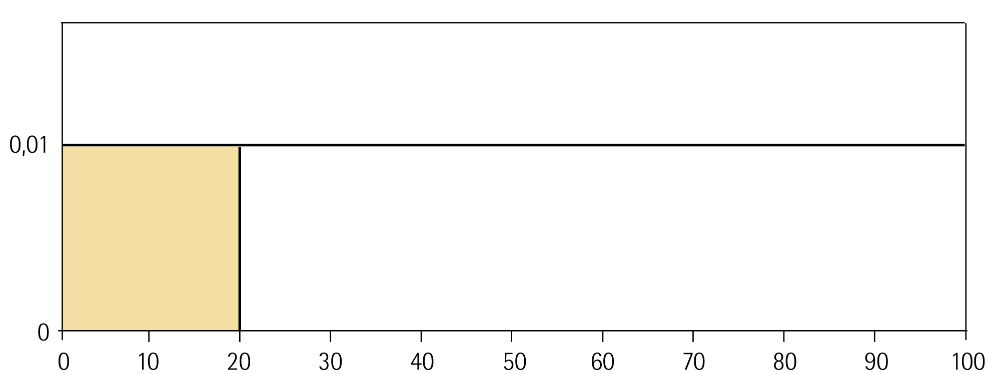
$P(x) = 0,1*x$

Se o primeiro convidado a se servir pegasse uma fatia de 20 cm, sua probabilidade de ganhar o prêmio seria:

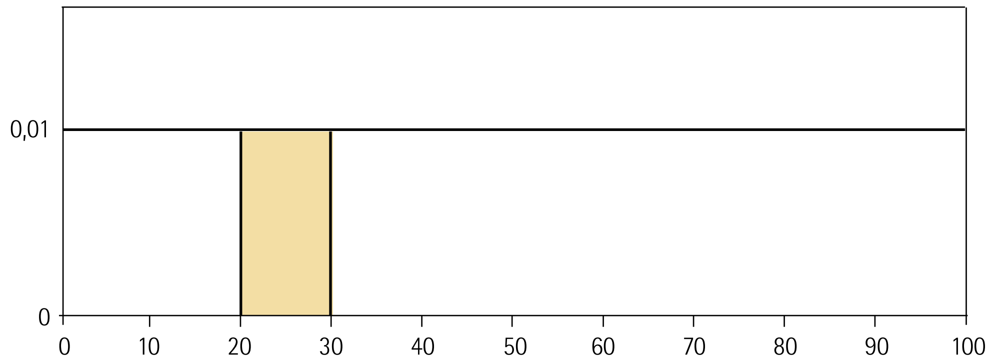
$P(20) = 0,1*20$

$P(20) = 0,2$

Para entender melhor o resultado, observe que uma fatia de 20 cm equivale a ter 20% do bolo. Formalmente, isso corresponde à área destacada no gráfico, conforme ilustramos a seguir:



Suponhamos agora que o segundo convidado pegasse uma fatia de 10 cm. A probabilidade de ele ser sorteado seria de 10%, conforme ilustrado na figura a seguir.



Assim, vimos que para distribuições contínuas, para determinar a probabilidade desejada deve-se obter a área do gráfico da distribuição.

No caso da distribuição uniforme, as figuras para as quais queremos calcular a área são retangulares, o que permite realizar a conta diretamente. No entanto, para outras distribuições, o cálculo é bastante complexo e essa determinação se faz por meio de valores tabelados.

7.8 Distribuição normal de probabilidades

A distribuição normal, também conhecida como distribuição gaussiana, é uma das mais importantes no estudo da estatística. Vamos agora estudar suas características e ver como utilizá-la para determinar probabilidades.

7.8.1 Condições de validade

Em termos gerais, dizemos que a distribuição normal descreve situações em que as características a seguir são observadas:

- Variável aleatória em questão é contínua.
- A probabilidade de ocorrência dos eventos possíveis depende somente da média e do desvio-padrão do conjunto.

Lembrete

Média: $\bar{x} = \dfrac{\sum x_i}{N}$

Variância: $\sigma^2 = \dfrac{\sum \left(x - \bar{x}\right)^2}{N}$

Desvio-padrão: $\sigma = \sqrt{\sigma^2}$

Nesses casos, temos uma distribuição de probabilidades, cuja forma geral traduz as seguintes características:

- Valor médio é o mais provável.
- Quanto mais longe da média, menos provável.
- O gráfico da distribuição tem um formato de sino.

Formalmente, a descrição das características necessárias ao conjunto são mais complexas, mas para o nível de discussão que cabe neste curso, elas são suficientes.

Apresentamos a seguir a expressão referente à curva gaussiana, bem como um gráfico em formato de sino, mas já adiantamos que essa expressão não será usada diretamente, pois sua utilização exige um domínio de matemática bastante avançado e fora do escopo deste curso. Na prática, utilizam-se valores tabelados, como veremos adiante.

Vale notar apenas que a função depende do valor médio e do desvio-padrão. A variável x refere-se ao valor da variável aleatória em estudo.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \; e \; \frac{-(x-\overline{x})^2}{2\sigma^2}$$

7.8.2 A variável z

Para a construção da tabela padronizada, primeiro se utiliza, em lugar do valor x da variável, o valor z, que nos diz quanto acima ou abaixo da média está o valor da variável em unidades de desvio-padrão.

O cálculo de z é dado por:

$$z = \frac{x-\overline{x}}{\sigma}$$

Note que, quando o valor de x é igual à média, o valor de z será zero.

Assim, se tivermos, por exemplo, um conjunto cuja média seja 50 e o desvio-padrão seja igual a 10:

- O valor x = 40 corresponderá a:

  - $z = \frac{40-50}{10} = \frac{-10}{10}$
  - $z = -1$

– Significado: o valor 40 está um desvio-padrão abaixo da média.

- Se x = 65, teremos:

– $z = \frac{65 - 50}{10} = \frac{15}{10}$

– z = 1,5

– Significado: o valor 65 está um desvio-padrão e meio acima da média.

- Para x = 50, que é a própria média, teremos:

– $z = \frac{50 - 50}{10} = \frac{0}{10}$

– z = 0

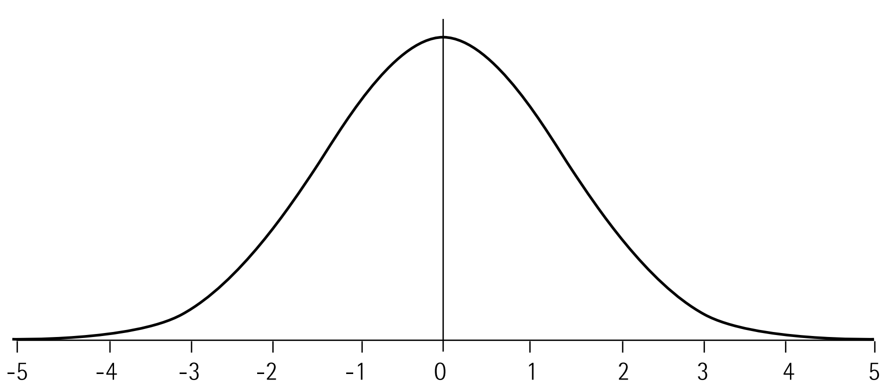
– Significado: o valor 50 não está nem acima nem abaixo da média.

Efetuando-se a substituição de x por z no cálculo da distribuição de probabilidades, teremos a função de distribuição dada por:

$$f(x) = \frac{1}{\sqrt{2\pi}}\ e\ \frac{-z^2}{2}$$

O gráfico da função anterior terá um máximo quando o valor de x for igual à média, portanto, z = 0. À medida que o valor de x se afasta deste, a função vai decaindo, sem nunca chegar a anular-se.

O gráfico a seguir explicita essa descrição.



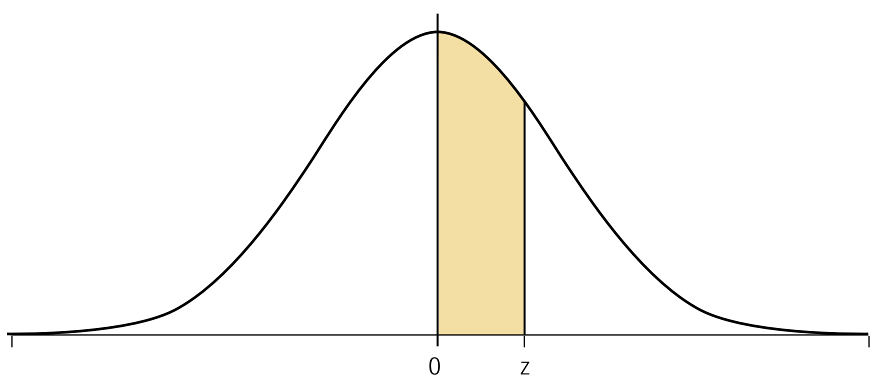
Assim como no exemplo para a distribuição uniforme, as probabilidades de se encontrar o valor da variável dentro de certo intervalo corresponderá à área do gráfico localizada no interior dos limites desse intervalo. Como a probabilidade de a variável ter qualquer valor possível é de 100%, a área sob a curva de toda a figura tem de ter valor 1.

No entanto, as "fatias" de probabilidade têm agora um tamanho difícil de determinar, pois esse "bolo" foi feito em forma de sino. Para que essa determinação seja possível, é preciso usar valores tabelados.

7.8.3 Distribuição normal padronizada

Para calcular as áreas, utilizaremos a tabela que nos dá a área, sob a curva gaussiana, padronizada em razão do valor de z. Assim, se temos um dado valor de x, precisamos inicialmente calcular o valor correspondente de z, a partir dos valores da média e do desvio-padrão, conforme vimos na seção anterior.

A tabela apresentada na pág. 115 nos trará a área correspondente ao intervalo entre 0 e z, conforme ilustrado na figura a seguir. Veremos adiante como determinar as probabilidades de intervalos diferentes deste a partir dos valores tabelados.



Leitura da tabela

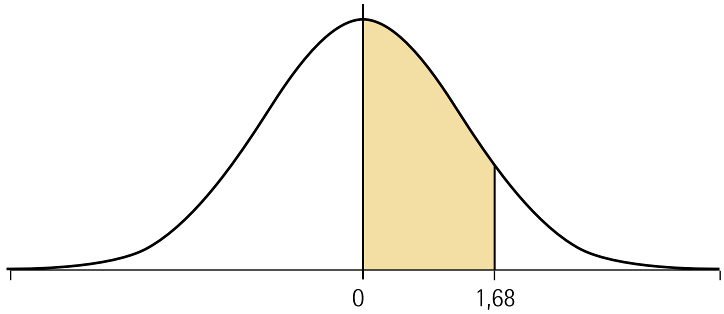
Na tabela citada há a informação sobre a área subentendida pela curva normal reduzida de 0 a z.

Dado um valor de z, devemos procurar na tabela qual o valor da área que corresponde a ele da seguinte forma:

- Na primeira coluna da tabela, o valor de z é dado até sua primeira casa decimal.
- Na primeira linha da tabela colocam-se os valores referentes à segunda casa decimal.
- O valor da área em questão será aquele que esteja na linha e coluna correspondentes.

Vejamos alguns exemplos:

- Área para z = 1,68.



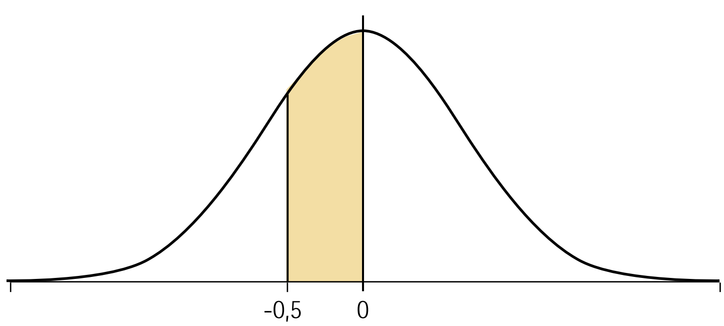
– Procure a linha cuja primeira coluna tenha o valor 1,6, a qual reproduzimos a seguir para facilitar a compreensão:

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545

– Nessa linha, procure a coluna correspondente à segunda casa decimal, ou seja, a coluna indicada pelo 0,08.

– A área indicada na figura, compreendida pela fatia que vai do z = 0 ao z = 1,68 será, portanto, A = 0,4535.

- Área para z = – 0,5



– Procure a linha cuja primeira coluna tenha o valor 0,5:

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,5	0,1915	0,195	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,219	0,2224

– Lembrando que 0,5 = 0,50, procure nessa linha a coluna correspondente à segunda casa decimal, ou seja, a coluna indicada pelo 0.

– A área compreendida pela fatia que vai do z = 0 ao z = 0,5 será, portanto, A = 0,1915.

Cálculo das probabilidades

Veremos agora diversos exemplos de como utilizar a tabela para determinar probabilidades que sigam a distribuição normal.

Uma empresa que realiza manutenção de computadores atende em média 25 clientes por dia, com desvio-padrão de 4,6.

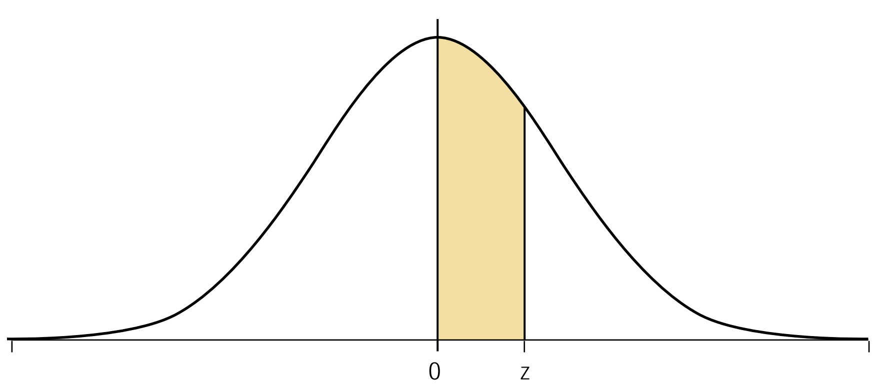
Em primeiro lugar, notamos que esse caso é passível de utilização da distribuição normal, pois:

- A probabilidade de atender um número acima ou abaixo da média é, em princípio, a mesma.
- Embora a variável "número de clientes" seja discreta, o parâmetro de análise é a variável z. Como a média e o desvio-padrão podem assumir qualquer valor, a divisão que leva à obtenção de z também pode ter como resultado qualquer valor real, portanto, a variável em análise é contínua.

Passamos, então, a analisar as probabilidades.

1. Qual a probabilidade de termos entre 25 e 28 clientes num dia?

- Identificação da área:
  - Valor mínimo: 25
  - Valor máximo: 28
- Cálculo de z:
  - $x_1 = 25$: $z = \dfrac{25-25}{4,6}$, logo, $z_1 = 0$
  - $x_2 = 28$: $z = \dfrac{28-25}{4,6}$, logo, $z_2 = 0,65$
- Identificação gráfica da área:



- Determinar o valor da área:
  - $z_1 = 0$, logo, o valor da área será o valor dado na tabela para z_2
  - Área para z = 0,65: A(0,65) = 0,2422

- Determinar o valor da probabilidade:

  – A probabilidade é igual à área dada, logo, P = 0,2422

2. Qual a probabilidade de termos entre 23 e 25 clientes num dia?

- Identificação da área:

  – Valor mínimo: 23

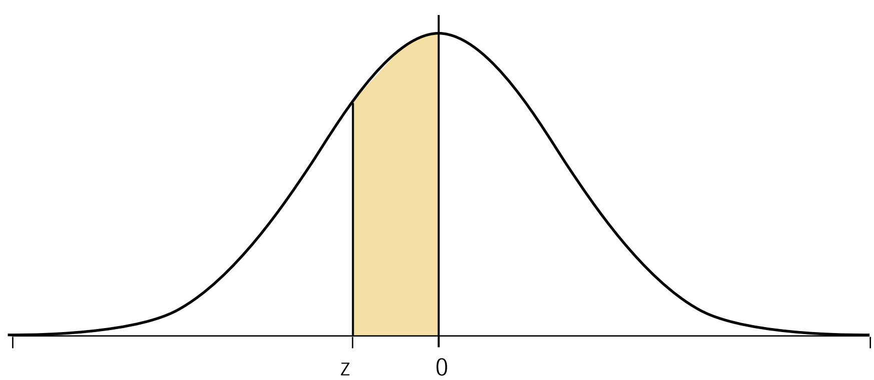
  – Valor máximo: 25

- Cálculo de z:

  – $x_1 = 23$: $z = \dfrac{23 - 25}{4,6}$, logo, $z_2 = -0,43$

  – $x_2 = 25$, logo $z_1 = 0$

- Identificação gráfica da área:



- Determinar o valor da área:

  – Por simetria, a área compreendida entre z = -0,43 e z = 0 é igual à área entre z = 0 e z = 0,43 dada na tabela.

  – Área para z = -0,43: A(0,43) = 0,1664

- Determinar o valor da probabilidade:

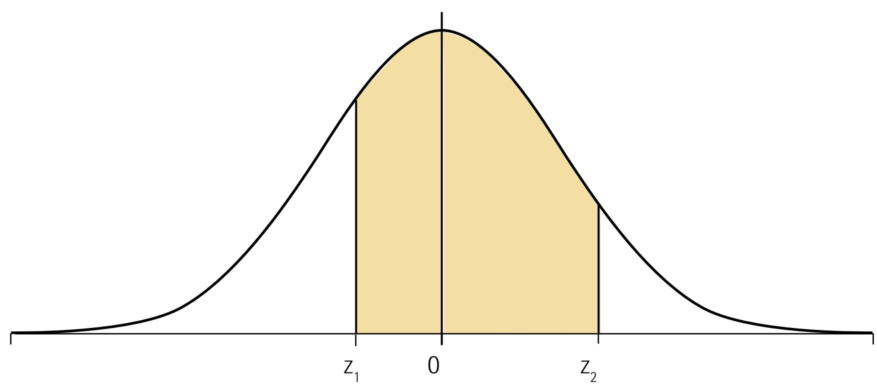
  – A probabilidade é igual à área dada, logo, P = 0,1664

3. Qual a probabilidade de termos entre 23 e 28 clientes num dia?

- Identificação da área:
  - Valor mínimo: 23
  - Valor máximo: 28

- Cálculo de z:
  - $x_1 = 23$: $z = \dfrac{23-25}{4,6}$, logo, $z_2 = -0,43$
  - $x_2 = 28$: $z = \dfrac{28-25}{4,6}$, logo, $z_1 = 0,65$

- Identificação gráfica da área:



- Determinar o valor da área:
  - A área total é a soma das áreas correspondentes aos valores de z_1 e z_2.
  - Área para z = -0,43: A(0,43) = 0,1664
  - Área para z = 0,65: A(0,65) = 0,2422
  - Área indicada: A = A(0,43) + A(0,65) = 0,1664 + 0,2422 = 0,4086

- Determinar o valor da probabilidade:
  - A probabilidade é igual à área dada, logo, P = 0,4086

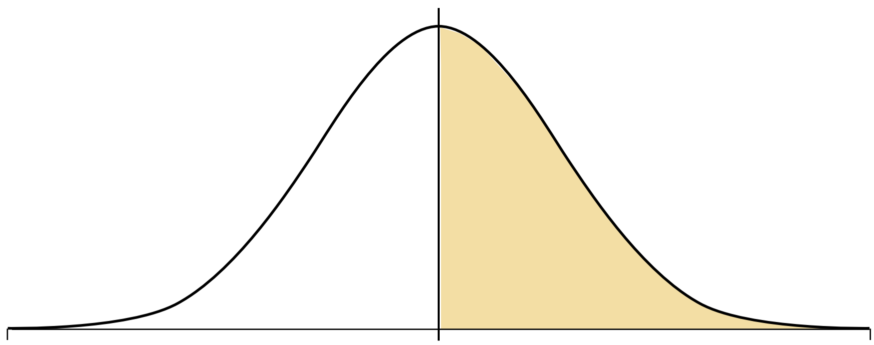
4. Qual a probabilidade de termos mais de 25 clientes num dia?

- Identificação da área:
  - Valor mínimo: 25
  - Valor máximo: infinito
  - Cálculo de z:

– $x_1 = 25$, logo, $z_2 = 0$

– $x_2 = \infty$, logo, $z_1 = \infty$

- Identificação gráfica da área:



- Determinar o valor da área:

  – A curva é simétrica, logo, o valor da área indicada corresponde à metade da área total da curva.

  – Como a área total é 1, A = 0,5

- Determinar o valor da probabilidade:

  – A probabilidade é igual à área dada, logo, P = A = 0,5

5. Qual a probabilidade de termos entre 26 e 30 clientes num dia?

- Identificação da área:

  – Valor mínimo: 26

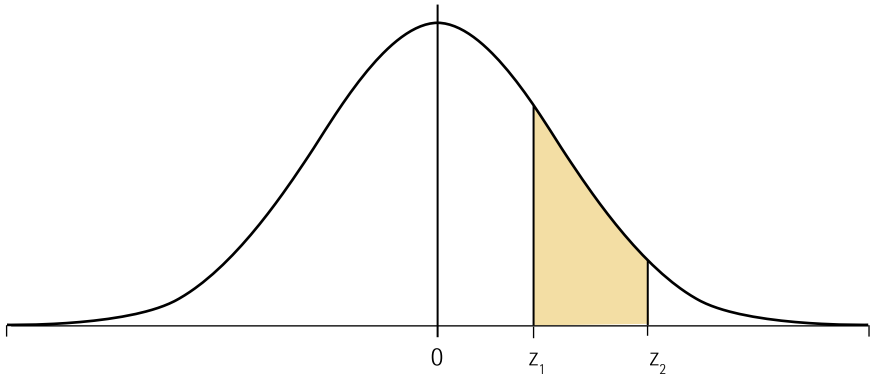
  – Valor máximo: 30

- Cálculo de z:

  – $x_1 = 26$: $z = \dfrac{26 - 25}{4{,}6}$, logo, $z_2 = 0{,}21$

  – $x_2 = 30$: $z = \dfrac{30 - 25}{4{,}6}$, logo, $z_1 = 1{,}09$

- Identificação gráfica da área:



- Determinar o valor da área:
  - Para encontrar a área total, é preciso descontar da área cujo limite superior é z_2 a área delimitada por z_1.
  - Área para z = 0,21: A(0,21) = 0,0832
  - Área para z = 1,09: A(1,09) = 0,3621
  - Área indicada: A = A(1,09) - A(0,21) = 0,3621 - 0,0832 = 0,2789

- Determinar o valor da probabilidade:
  - A probabilidade é igual à área dada, logo, P = 0,2789.

Tabela

Área subentendida pela curva normal reduzida de 0 a z.

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0	0,004	0,008	0,012	0,016	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,091	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,148	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,17	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,195	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,219	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,258	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,291	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,334	0,3365	0,3389
1	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,377	0,379	0,381	0,383
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,398	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177

1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,437	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,475	0,4756	0,4761	0,4767
2	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,483	0,4834	0,4838	0,4842	0,4846	0,485	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,489
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,492	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,494	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,496	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,497	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,498	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,499	0,499
3,1	0,499	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5

8 INFERÊNCIA E ESTIMAÇÃO

Como tópicos finais do curso, passaremos agora a discutir dois temas de grande interesse na área da estatística.

Num primeiro bloco trataremos da questão do uso de amostras e da confiabilidade dos resultados obtidos a partir de dados amostrais.

Em seguida discutiremos brevemente como métodos estatísticos verificam haver, ou não, relações entre diferentes grandezas e o significado da existência de correlações.

8.1 Definições

Basta um pequeno gole de suco para saber que gosto ele tem, mas alguns grãos de feijão não lhe permitirão conhecer todo o sabor de uma feijoada. Nesses dois casos, a questão é podermos conhecer o todo a partir de uma parcela.

A possibilidade de se estudar o todo a partir do conhecimento de uma parte só é absolutamente crucial quando queremos conhecer as características de um conjunto muito grande. Imagine, por exemplo, como seria impraticável ter índices de inflação, por menores que fossem, se tivéssemos que saber todos os preços de todos os produtos em todas as lojas do país. No entanto, se a escolha da amostra não for benfeita, poderemos ter uma descrição que não condiga com os fatos.

Há diversos métodos estatísticos que buscam garantir e analisar o grau de confiabilidade nesses casos. Uma discussão aprofundada do assunto é bastante complexa e não cabe no âmbito deste curso, assim, em muitos momentos, vamos tratar as questões referentes à amostragem de maneira puramente qualitativa.

8.1.1 População e amostra

Chamamos de população um conjunto de indivíduos com pelo menos uma característica observável. Por exemplo, podemos ter como população o conjunto de preços praticados no varejo para um dado produto.

Uma população pode ser considerada finita ou infinita, dependendo da existência de uma limitação da quantidade de elementos de que é composta.

Dizemos que a população é finita quando o número de indivíduos é limitado. Temos como exemplo disso o número de alunos de uma faculdade, a quantidade de indústrias em uma cidade etc.

Chamamos de infinita uma população com número ilimitado de elementos ou, mais especificamente, com populações cujos elementos são processos que geram itens. Um exemplo disso seria o jogo sucessivo de um dado, pois podemos continuar indefinidamente fazendo novas jogadas se for de interesse, o que dá um caráter ilimitado ao número de jogadas.

O estudo de uma população pode ser feito a partir da obtenção de dados de apenas parte dos elementos dela. Chamamos de amostra a esse subconjunto da população.

Uma amostra tem por objetivo representar uma população de tal modo, que as informações a respeito dela sejam significativas para toda a população. Assim, as escolhas de amostras devem ser feitas de maneira criteriosa e de acordo com as características da população.

8.1.2 Censo e amostragem

Para obtermos informações a respeito de uma população há duas possibilidades: podemos averiguar a informação para todos os elementos da população ou podemos escolher apenas parte do conjunto para tal averiguação.

Quando o universo estudado inclui toda a população, temos a realização de um censo.

Quando o universo é subconjunto finito da população, então temos um caso de estudo por amostragem.

A utilização de censo ou amostragem num estudo depende de uma série de fatores. O mais usual é utilizar amostras por causa da dificuldade ou, mesmo, impossibilidade de se estudarem todos os componentes da população.

Vantagens do censo

A utilização do censo tem como principal vantagem a possibilidade de termos precisão completa nos resultados. Quando isso é necessário, a única abordagem possível é o censo. No entanto, é importante notar que, na prática, há poucos casos que exigem estudos estatísticos em que tal precisão é requerida.

Uma segunda situação em que a utilização de censo é mais indicada é aquela em que uma população é tão pequena que o custo e o tempo gastos com a realização de um censo são pouco maiores que para uma amostra. Por exemplo, se quisermos estudar alguma característica referente a uma determinada turma de um curso, é mais trabalhoso estudar as maneiras de montar uma amostra apropriada que obter diretamente os dados de todos os alunos.

De forma similar, se o tamanho da amostra é grande, em relação ao da população, o esforço adicional requerido por um censo pode ser pequeno. Com relação ao tamanho da amostra e à precisão do estudo, vamos discutir esse assunto detalhadamente mais adiante.

Vantagens da amostra

Na maior parte dos casos, revela-se mais vantajosa a utilização da amostra que a utilização do censo.

A primazia da amostragem é evidente em casos em que são necessários testes destrutivos. Um exemplo claro disso são os estudos de qualidade de um produto quanto à resistência, à durabilidade etc. Quase anedótico seria pensar em um estudo sobre a qualidade de produção de fósforos em que um censo implicaria, necessariamente, na destruição de toda a produção.

Também é melhor utilizar a amostragem quando estudamos alguma informação que varia rapidamente, pois a obtenção de uma grande quantidade de dados pode demorar mais do que o tempo em que a informação permanece estável. Imagine, por exemplo, um levantamento de preços que levasse em conta todos os estabelecimentos comerciais do país. Até que fossem coletados os dados, os preços já teriam sido reajustados diversas vezes e a informação recolhida não teria mais significado.

Outra razão para escolher o estudo de amostras é o custo, pois em diversos casos os gastos necessários para a realização de um censo podem ser proibitivos.

Dissemos na seção anterior que o censo traz a possibilidade de precisão absoluta. No entanto, na prática, a realização de um censo pode ter uma complexidade tal que pode acarretar erros durante a coleta dos dados. Imagine o censo realizado pelo IBGE, em que há milhares de recenseadores buscando coletar informações a respeito de milhões de pessoas. É fácil imaginar a impossibilidade de ocorrer tal coleta de dados sem que haja erros.

Outra razão de escolha de estudo por amostragem diz respeito à variedade e profundidade de informações que se busca. Por questões de custo e tempo, especialmente, um censo limita a pesquisa a poucas características por item.

8.2 Técnicas de amostragem

Existem diversas técnicas para a obtenção de amostras. Citamos aqui algumas das principais características das classes de técnicas existentes.

8.2.1 Classificação

Podemos classificar as amostras em dois grandes grupos: o das probabilísticas e o das não probabilísticas.

Probabilísticas

As amostras probabilísticas são aquelas em que a escolha dos elementos é feita de forma aleatória, utilizando-se alguma maneira para sortear os elementos. Essa técnica é segura para inferir características da população e é a mais comumente utilizada nos estudos de caráter econômico e social.

Não probabilísticas

As técnicas não probabilísticas utilizam escolha deliberada dos elementos e nem sempre permitem generalizar os resultados para a população, pois essa opção pode enfatizar ou atenuar alguma característica. Essas técnicas são bastante utilizadas em estudos médicos, por exemplo.

8.2.2 Duas técnicas probabilísticas

Amostragem aleatória simples

Essa técnica é mais elementar é frequentemente utilizada e consiste em sortear elementos da população de maneira completamente aleatória.

Como todos os elementos da população têm a mesma probabilidade de ser selecionados, deduz-se que a amostra tem alta probabilidade de ter composição representativa das proporções encontradas na população.

Por exemplo, se queremos estudar a opinião dos alunos de uma faculdade a respeito do lançamento de um produto, podemos sortear alunos para responder a um questionário. Se a faculdade tiver um grande número de mulheres, é bastante provável que sejam sorteadas mais mulheres que homens, e assim as características da população de alunos fica representada na amostra.

Amostragem estratificada

A amostragem estratificada realiza também sorteios para a montagem da amostra, mas considera as diferenças relevantes entre as características dos subconjuntos, ou seja, de seus diversos estratos.

No caso em que temos diferenças significativas e queremos ter certeza de que cada grupo está bem representado, o sorteio dos elementos leva em consideração tais divisões.

Por exemplo, uma população é composta de 20% de mulheres e 80% de homens e quero fazer uma amostra composta de 100 pessoas. Se for relevante ter uma proporção correta de homens e mulheres no estudo, sorteamos 20 mulheres e 80 homens para compor a amostra.

8.2.3 Duas técnicas não probabilísticas (não aleatórias)

Muito embora as técnicas probabilísticas nos permitam realizar estudos mais confiáveis, em geral há situações em que amostras não probabilísticas se mostram mais interessantes ou viáveis.

Amostragem acidental

Em algumas ocasiões, as amostras são formadas por elementos escolhidos somente por causa de sua disponibilidade.

Essa forma de escolha é muito comum em pesquisas de opinião, em que os entrevistados são acidentalmente escolhidos ao passarem por uma praça movimentada, por um grande supermercado ou por outro lugar qualquer em que entrevistadores são colocados para selecionar as pessoas aleatoriamente.

Também é comum essa forma de amostragem no caso de estudos de novos medicamentos, quando se constrói a amostra em razão de haver pessoas que buscam tratamento para a doença para a qual o remédio foi desenvolvido e que se disponham a tomá-lo apesar de ainda não ser completamente conhecido.

Amostragem intencional

Por fim, podemos ter ainda a escolha intencional, segundo um critério preestabelecido. Essa técnica é indicada quando se tem interesse apenas em informações sobre um conjunto restrito ou quando se avalia que as opiniões de certos indivíduos sejam mais significativas.

Por exemplo, um gerente-geral pode considerar que a opinião de certo conjunto de funcionários antigos, a respeito de uma alteração na política da empresa, seja mais significativa que a opinião da totalidade dos funcionários, por causa do conhecimento que já adquiriram sobre seu local de trabalho.

8.3 Representatividade da amostra

Discutimos anteriormente alguns critérios qualitativos para escolha de amostra. Agora, é preciso analisar a confiabilidade deles no que diz respeito a inferir características da população como um todo a partir dos dados desse subconjunto.

Por exemplo, vemos em pesquisas de intenção de voto que o resultado tem certa margem de erro. Se essa margem for de 3%, sabemos que se dois candidatos têm menos do que esse percentual de diferença, eles estão na verdade empatados, ou se um candidato teve uma variação de 2% na proporção de eleitores que dizem pretender votar nele, pode ser que na verdade ele tenha se mantido estável. Considerar, nessas condições, que somente o número, sem ver qual sua incerteza, pode trazer análises incorretas da realidade, e, se for preciso tomar decisões em que esses dados sejam relevantes, um erro de análise desse tipo pode levar a decisões também errôneas.

Em suma, quando realizamos um estudo por amostragem, a grande questão que se coloca é: quão representativa é certa amostra?

A determinação precisa dessas incertezas depende, muitas vezes, de conhecimentos de estatística que vão além da proposta deste curso, no entanto é crucial ter noções de como essas análises são feitas.

Para mostrar qual tipo de procedimento se utiliza nesse tipo de análise, vamos ver como se determina o erro de uma média amostral, ou seja, como se compara a média obtida por amostragem com a média real da população quando conhecemos o desvio-padrão para a população.

8.4 Distribuições amostrais

Para uma mesma população, podemos selecionar uma amostra com um número menor de elementos de diversas maneiras. O conjunto de todas as amostras possíveis com um determinado número de elementos é chamado de distribuição amostral.

Para que fiquem claros os critérios de comparação, vamos ver como se comparam a média real e as médias de cada uma das amostras em um caso em que todos, população e amostra, são conhecidos.

8.4.1 Nova notação

Como agora teremos valores de média e desvio referentes a cada amostra, ao conjunto das amostras e à população, é preciso denotá-las de maneira diferente:

- **População**
  - – O símbolo a ser utilizado a partir de agora para a média da população será a letra grega μ.
  - – O símbolo a ser utilizado no desvio-padrão da população será a letra grega σ.
- **Amostra específica**
  - – O símbolo para uma média específica será $\overline{x}$.
  - – Não discutiremos aqui os desvios de amostras específicas.

- **Distribuição amostral**

  – A média do conjunto de amostras será indicada por $\mu_{\bar{x}}$.

  – O símbolo $\sigma_{\bar{x}}$ denotará o desvio-padrão da distribuição amostral.

O objetivo agora será mostrar como se comparam a média e o desvio-padrão de uma distribuição amostral, $\mu_{\bar{x}}$ **e** $\sigma_{\bar{x}}$, com os respectivos valores para a população, μ e σ, que são os valores reais**.**

Exemplo

Vamos ver em detalhes um exemplo em que temos os dados referentes a gasto com roupas de uma população de 5 pessoas e analisaremos a média que seria obtida para as amostras compostas de 2 elementos.

Ressalto que um estudo por amostragem, nesse caso, não teria razão de ser, mas o objetivo aqui não é obter informações a respeito do conjunto, mas sim mostrar didaticamente como se comparam a média de uma população com as médias de suas amostras e, para efetuar tal comparação, é necessário que conheçamos todos os dados.

Dados da população

Na tabela a seguir são apresentados os gastos das 5 pessoas que fazem parte da população a ser analisada.

Pessoa	A	B	C	D	E
Valor	100	130	150	170	200

A partir desses valores, calculamos a média e o desvio para população, seguindo o procedimento já estudado. Esses são os valores que nos servirão de parâmetro na comparação com as médias amostrais:

$\mu = 150$ e $\sigma = 38$

A partir de agora, vamos analisar as médias encontradas caso tenhamos amostras compostas de somente 2 elementos.

Construção das amostras

Existem 10 maneiras de se escolher 2 elementos entre 5, pois este é um caso de combinação, onde se escolhem 2 elementos entre 5 e a ordem de escolha não é importante. Logo, o número possível de amostras será:

$$N = \frac{5!}{2!(5-2)!} = \frac{5 * 4}{2} = 10$$

As amostras possíveis serão: AB, AC, AD, AE, BC, BD, BE, CD, CE e DE.

Média e desvio de cada amostra

Na tabela a seguir colocamos os valores individuais e a média para cada uma das dez amostras.

	Amostra	Gastos (x)		Média ($\bar{x}$)
1	AB	100	130	115
2	AC	100	150	125
3	AD	100	170	135
4	AE	100	200	150
5	BC	130	150	140
6	BD	130	170	150
7	BE	130	200	165
8	CD	150	170	160
9	CE	150	200	175
10	DE	170	200	185

Note, em primeiro lugar, que os valores mais próximos da média real, digamos até R$ 10,00 acima ou abaixo da média, são mais frequentes que os valores mais distantes, uma vez que na própria população é mais fácil encontrarmos os valores similares à média real.

Se fizermos a média das médias das amostras teremos:

$$\mu_{\bar{x}} = \frac{130+150+170+200+150+170+200+170+200+200}{10}$$

$$\mu_{\bar{x}} = \frac{1500}{10} = 150$$

Esse valor é chamado de média da distribuição amostral, pois é a média dos valores encontrados para as diversas amostras possíveis. Note que esse valor é igual à média da população.

O fato de termos a média da distribuição igual à média da população é esperado, visto que cada elemento aparece um mesmo número de vezes nas amostras. Nesse caso, cada elemento apareceu 4 vezes. Assim, a distribuição amostral terá contribuição equivalente de todos os elementos da população e, com isso, mesma média.

No que diz respeito ao desvio-padrão, no entanto, essa igualdade não se observa. O desvio-padrão da população é 38 e o desvio-padrão das médias da população é 22.

Lembrando que o desvio-padrão é uma medida de dispersão, que reflete a diversidade de valores, podemos entender tal diferença pelo fato de uma amostra ter diversidade menor que uma população e, desse modo, os valores referentes à dispersão tenderão a ser menores também.

8.4.2 Comparação entre distribuição amostral e população

Vejamos agora o caso geral em que temos uma distribuição amostral para amostras com n elementos.

Assim como no exemplo estudado, a média da distribuição é igual à média da população e o desvio-padrão tende a ser menor que o desvio referente à população. Em estudos reais, em que há um grande número de dados, a distribuição dos valores das médias tende a seguir a curva normal, com distribuição em torno da média da população.

A média da distribuição será, portanto:

$$\mu_{\bar{x}} = \mu$$

A dispersão da distribuição será tanto maior quanto maior for a dispersão na população de onde as amostras são extraídas e tanto menor quanto maior for a amostra, seguindo a relação a seguir:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

8.5 Estimação

Na seção anterior vimos como se distribuem os valores das médias referentes a todas as possibilidades de amostras de um dado tamanho n. Nesta seção discutiremos como estimar a média da população a partir da média de uma amostra.

Em razão dos objetivos desta disciplina, não cabe realizar uma discussão quantitativa extensa deste tema, portanto escolhemos abordar somente o caso em que o desvio-padrão da população é conhecido e em que a incerteza do valor seja igual para valores acima ou abaixo da média encontrada.

Note que o objetivo desta seção é apenas fazer com que você, aluno conheça o tipo de procedimento que nos permite estimar a confiabilidade de um valor obtido por amostragem, sem esperarmos que ele domine de fato a execução do procedimento descrito.

8.5.1 Intervalos de confiança bilateral para σ conhecido

Suponhamos um estudo amostral em que foi obtida uma média x representando uma população cujo valor médio é μ.

A diferença entre o valor medido e o valor real será chamado de "erro", o qual será denotado pela letra e. Vamos analisar a probabilidade de termos um determinado erro máximo.

Para certo erro, e, estima-se que o valor real esteja dentro do intervalo:

$$\mu = \overline{x} - e$$

No qual e é o erro estimado dado por:

$$e = \frac{z.\sigma}{\sqrt{n}}$$

O valor de z é a distância da média em unidades de desvio-padrão, conforme já discutido, para a distribuição normal padronizada. Como essa grandeza serve como referência para a determinação das áreas sob a curva gaussiana, e, portanto, para a determinação da probabilidade de intervalos de valores da grandeza estudada, esse parâmetro nos dará o grau de confiança que se pode ter no valor medido para a média.

Podemos utilizar essa expressão para estimarmos o erro de uma média com certa probabilidade de acerto.

Essa estimação envolve determinar o valor de z correspondente a certa probabilidade. Já havíamos visto como determinar a probabilidade a partir de z, agora, vamos trilhar o caminho oposto. Vejamos nos exemplos a seguir como se dá na prática tal determinação.

Determinação de z para uma dada probabilidade

Vimos anteriormente como determinar a probabilidade a partir de valores para z. Agora, utilizaremos a mesma tabela para encontrar o valor de z que corresponde a certa probabilidade.

Digamos que eu queira, por exemplo, estimar uma média com 90% de certeza. Para isso, em primeiro lugar, tenho que encontrar o z que corresponde a tal probabilidade, ou seja, o valor de z que delimita uma região da curva normal com 0,45 para cada lado em torno da média.

Na prática, isso significa encontrar na tabela da distribuição normal padronizada o número mais próximo a 0,45.

Conforme vemos na reprodução da linha 1,6 da tabela citada, o valor mais próximo àquele que procuramos é o 0,4505, que corresponde a termos z = 1,65.

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545

Dessa maneira, verificamos que 90% dos resultados possíveis estão localizados na região em que $-1{,}65 < z < 1{,}65$.

Fazendo o mesmo para a probabilidade de 80%, procuraremos o valor mais próximo a 0,40 na tabela. Conforme vemos a seguir, o valor mais próximo é o 0,4015, que corresponde a z = 1,29.

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,398	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177

O procedimento anterior é utilizado para a escolha de qualquer intervalo de confiança, no entanto, é usual utilizar confianças de 90%, 95% ou 99%, que correspondem, respectivamente, a z = 1,65, z = 1,96 e z = 2,58.

Intervalo de confiança

Uma vez encontrado o z, utilizamos a expressão dada anteriormente para determinar o erro associado à média:

$$e = \frac{z.\sigma}{\sqrt{n}}$$

Assim, se tivermos, por exemplo, uma média $\bar{x} = 24{,}2$, obtida a partir de uma amostra com n = 36, numa população com desvio-padrão $\sigma = 3$, estimamos o erro da seguinte maneira:

Em primeiro lugar, substituímos os valores de n e σ na expressão:

$$e = \frac{z.3}{\sqrt{36}} = \frac{z.3}{36} = \frac{z}{2}$$

Se quisermos saber com 90% de confiança qual a média da população, constaremos que z = 1,65 e, portanto, o erro será de 0,825.

O valor real tem 90% de probabilidade de estar entre o valor da média mais o valor do erro e o valor da média menos o valor do erro:

- Limite inferior do intervalo: 24,2 - 0,825 = 23,375.
- Limite inferior do intervalo: 24,2 + 0,825 = 25,025.

Em suma, a média da população tem 90% de probabilidade de estar entre os valores 23,375 e 25,025.

Se quisermos uma maior certeza de que o valor real esteja no intervalo, realizamos o procedimento descrito para o percentual desejado e utilizamos o valor correspondente de z para encontrar o valor do erro nesse caso.

Na tabela a seguir, apresentamos os resultados para confianças desejadas de 90%, 95% e 99%:

Confiança desejada	z	e	Intervalo
90%	1,65	0,825	23,375 a 25,025
95%	1,96	0,980	23,220 a 25,180
99%	2,58	1,290	23,110 a 25,690

Ou seja, encontrando a média 24,2 nas condições citadas, sabemos que:

- Há 90% de probabilidade de o valor real estar ente 23,375 e 25,025.
- Há 95% de probabilidade de o valor real estar ente 23,220 e 25,180.
- Há 99% de probabilidade de o valor real estar ente 23,110 e 25,690.

8.5.2 Observações finais

Vimos aqui algumas características principais das análises que se devem realizar quando fazemos um estudo por amostragem.

Abordamos somente um caso mais simples, em que se busca a média de uma população com desvio-padrão conhecido e estimamos a incerteza no valor da média considerando incerteza para valores acima ou abaixo daquele encontrado.

Estudos reais exigem cuidados muito mais abrangentes e minuciosos, porém, esses detalhamentos e aprofundamentos seguem todos uma metodologia cujas características básicas estão presentes nas discussões apresentadas.

Por exemplo, quando o desvio-padrão é desconhecido, utiliza-se um procedimento similar ao descrito anteriormente, mas a estimativa dessa incerteza não utiliza a distribuição normal, mas sim outra chamada de distribuição de *student*. Nesses casos, há alguns parâmetros adicionais a considerar, mas o importante aqui é saber que ela nos dará uma incerteza maior nos valores, refletindo o fato de termos menos informações a respeito da população.

Além da média, outras grandezas, como a proporção de cada tipo de dado, podem ser obtidas utilizando amostras. Também para essas grandezas temos técnicas de análise de confiabilidade similares às expostas para a média.

O ponto crucial, entretanto, é saber que estudos estatísticos a partir de amostras permitem de fato a obtenção de dados a respeito da população e que existem técnicas que nos permitem conhecer os níveis de confiabilidade dos valores assim obtidos. Também é importante ressaltar que sempre existe a probabilidade de o valor real não estar dentro da margem de erro estimada, mas que, na maioria das vezes essa probabilidade é muito pequena e os resultados podem ser utilizados como parâmetro para a compreensão do conjunto.

Exemplos de aplicação

1) Imagine uma situação em que seja mais indicado utilizar um censo em lugar de um estudo por amostragem. Para isso, escolha a grandeza a ser estudada e considere fatores como confiabilidade, custo, viabilidade etc.

2) Imagine uma situação em que seja mais indicado utilizar uma amostra para se estimar uma certa grandeza. Descreva a grandeza e a população escolhidas e justifique sua escolha.

   a) Considere uma população finita.

   b) Considere uma população infinita.

3) Imagine uma situação em que seja indicado utilizar uma técnica não probabilística de amostragem:

   a) Amostragem acidental.

   b) Amostragem intencional.

4) Para escrever uma reportagem a respeito de uma nova lei municipal de uma cidade com cerca de 10.000 habitantes, um jornalista resolve realizar uma pesquisa de opinião. Para cada abordagem a seguir, aponte o problema de ele escolher:

   a) Fazer um censo de opiniões na cidade.

   b) Entrevistar todos os alunos da escola ao lado.

   c) Entrevistar as primeiras 20 pessoas que passarem em frente ao jornal.

   d) Entrevistar somente pessoas da cidade vizinha.

5) Uma empresa tem 5 funcionários que realizam atendimento ao público. O número de atendimentos realizado por empregado em uma semana é listado na tabela a seguir:

Abel	Sílvia	Denise	Fábio	Gílson
72	54	51	63	60

a) Calcule a média e o desvio-padrão dessa população.

b) Determine quantas amostras de 3 pessoas são possíveis para esse universo.

c) Construa cada uma das amostras e calcule a média amostral de cada uma delas.

d) Calcule a média da distribuição e seu desvio-padrão.

e) Compare os valores da distribuição com os valores da população.

6) Uma empresa que presta assistência técnica realizou um estudo detalhado a respeito de atendimento de clientes, tendo verificado que a nota média de satisfação dos clientes era 7, com desvio-padrão 2. Buscando melhorar esse índice, foi contratada uma empresa de treinamento para aperfeiçoar o desempenho dos funcionários. Após o treinamento, foi feito um estudo por amostragem, utilizando 25 atendimentos, e obteve-se a nota 8. Utilizando o procedimento descrito no texto:

a) Construa os intervalos de confiança para a média encontrada após o treinamento.

b) Com base no resultado anterior, discuta se se pode considerar que o treinamento foi eficiente.

Resolução

1) Pergunta com resposta aberta. Deve-se escolher uma situação em que o número de elementos no conjunto universo seja pequeno ou algum caso em que seja realmente importante termos precisão absoluta.

2) Pergunta com resposta aberta.

a) Deve-se escolher uma situação que não exija precisão absoluta e em que seja pouco viável fazer uma tomada de dados que inclua todos os elementos.

b) Nos casos em que a população é infinita sempre é necessário utilizar amostras.

3) Pergunta com resposta aberta.

a) Deve-se escolher uma situação em que não seja prático ou viável fazer uma escolha aleatória.

b) Deve-se escolher uma situação em que haja algum subconjunto que, por si, seja mais representativo. Por exemplo, numa pesquisa de opinião a respeito do lançamento de um programa, perguntar a opinião somente de alguns técnicos experientes da área, escolhidos por causa de seu conhecimento.

4)

a) Não é viável perguntar a opinião de todos os moradores para fazer uma reportagem em um jornal.

b) O conjunto não seria representativo, por incluir somente pessoas de um mesmo grupo social.

c) O conjunto poderia não ser representativo, por ter um número muito pequeno de pessoas na amostra.

d) O conjunto não seria representativo, já que a amostra teria sido escolhida em um universo diferente daquele que se queria estudar.

5)

a) Média: 60, desvio-padrão: 8,2.

b) N = (5*4*3)/(3!) = 10

c)

Amostra		Atendimentos			Média
1	asd	72	54	51	59
2	asf	72	54	63	63
3	asg	72	54	60	62
4	adf	72	51	63	62
5	adg	72	51	60	61
6	afg	72	63	60	65
7	sdf	54	51	63	56
8	sdg	54	51	60	55
9	sfg	54	63	60	59
10	dfg	51	63	60	58

d) Média: 60, desvio-padrão: 3,2.

e) Conforme esperado, as médias são iguais. O desvio-padrão da distribuição amostral é menor, pois as amostras têm menos diversidade que a população. O valor não coincide com a divisão pela raiz de n porque a população e a amostra são pequenas e essa relação só é válida no limite de grandes conjuntos.

6)

a) Construa os intervalos de confiança para a média encontrada após o treinamento.

Confiança desejada	z	e	Limite mínimo	Limite máximo
90%	1,65	0,66	7,34	8,66
95%	1,96	0,78	7,22	8,78
99%	2,58	1,03	6,97	9,03

b) Sim, podemos considerar que o treinamento foi eficiente, visto que apenas quando a confiança exigida é de 99% é que vemos aparecer a probabilidade de a nota real ser igual à nota obtida antes do treinamento.

Resumo

Noções de probabilidade

1. Variáveis aleatórias são aquelas cujo resultado depende do acaso.

2. O estudo das probabilidades busca determinar quais dos resultados possíveis serão mais raros ou comuns.

3. Para calcular uma probabilidade, devemos dividir o número de eventos favoráveis ao aspecto escolhido pelo número total de eventos possíveis (ou espaço amostral).

4. Os valores para calcular probabilidades podem ser empíricos ou analíticos.

5. Uma probabilidade sempre será um número entre 0 e 1.

Análise combinatória

1. A contagem de eventos independentes associados se baseia no princípio da multiplicação.

2. Há algumas situações específicas de maior interesse.

3. Permuta sem repetição:

a) Troca de n elementos diferentes em n posições.

b) $P_n = n!$

4. Permuta com repetição:

a) Troca de n elementos em n posições, havendo elementos repetidos ou iguais.

b) $P_n^{a,b,c...} = \dfrac{n!}{a!b!c!...}$

5. Arranjo:

a) Maneiras possíveis de se ter n elementos em p posições quando a ordem é relevante.

b) $A_{n,p} = \dfrac{n!}{(n-p)!}$

6. Combinação:

a) Maneiras possíveis de se ter n elementos em p posições, quando a ordem não é relevante.

b) $C_{n,p} = \dfrac{n!}{p!(n-p)!}$

Distribuições de probabilidade

1. Trazem as probabilidades de cada evento possível do universo estudado.

2. Há distribuições discretas e contínuas.

3. A distribuição binomial é discreta e segue a expressão:

$$P(k) = \frac{n!}{k!(n-k!)} p^k q^{n-k}$$

4. A distribuição normal é contínua e para a obtenção da probabilidade referente a um intervalo utilizamos a tabela da distribuição normal padronizada.

Inferência e estimação

1. Há diversas situações em que um estudo estatístico é feito utilizando dados apenas a respeito de parte dos elementos do conjunto estudado.

2. Quando se obtém dados de todos os elementos, temos a realização de um censo; quando temos dados somente de parte dos elementos, temos um estudo por amostragem.

3. Para um mesmo conjunto universo há diversas possibilidades de escolha de amostra. O conjunto dos resultados possíveis para as diversas amostras é chamado de distribuição amostral.

4. Existem técnicas que nos permitem selecionar amostras de maneira tal que a escolhida seja representativa do conjunto.

5. Existem maneiras de se estimar qual o grau de confiabilidade de resultados utilizando amostras.

6. A média encontrada em uma pesquisa amostral não precisa ser idêntica à da população.

7. O valor para amostra será tanto mais confiável quanto maior o tamanho da amostra e menor o desvio-padrão (menor diversidade, menor dispersão).

Exercícios

Questão 1. O seu chefe precisa tomar uma decisão acerca da implantação de uma nova unidade e pediu para você fazer uma estimativa dos gastos com salários. Para tanto, você fez uma pesquisa com 225 trabalhadores da região em que será instalada a fábrica, seguindo a distribuição de cargos e funções, e chegou à média de R$ 1.950,00, com desvio-padrão de R$ 298,00. Baseado nesses dados, qual foi a estimativa apresentada com 95% de confiabilidade?

A) 1950 $\pm$ 79,6 reais

B) 1950 $\pm$ 43,4 reais

C) 1950 $\pm$ 38,9 reais

D) 1950 $\pm$ 31,4 reais

E) 1950 $\pm$ 28,0 reais

Resposta correta: alternativa C.

Análise da alternativa

A sequência de cálculos necessária é a seguinte:

Valor estimado = Valor provável $\pm$ erro esperado

- Valor provável = 1950

- Erro esperado = z_c • erro padrão = 1,96 • 19,9 = 38,9

– z_c = 1,96 (coeficiente para 95% de confiabilidade, vide tabela)

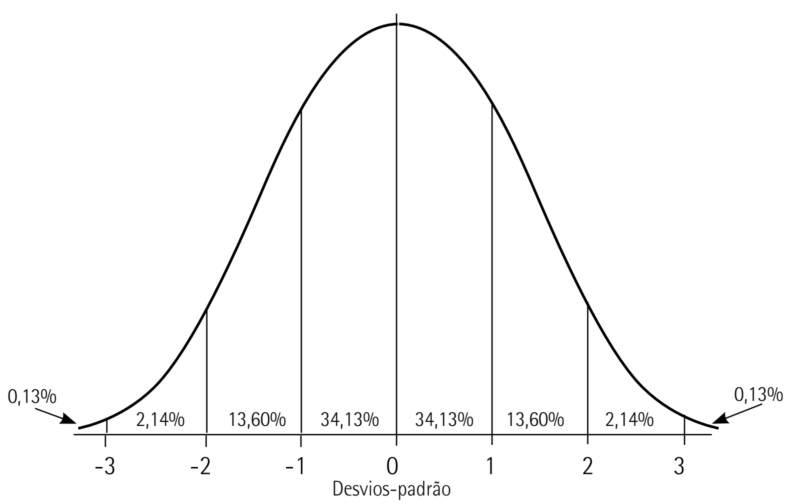
– Erro-padrão => $\sigma_x = \frac{298}{\sqrt{225}} = 19,9$

Valor estimado = 1950 ± 38,9 reais com 95% de confiabilidade.

Portanto, a alternativa que apresenta o valor correto é a C.

Questão 2. (ENADE) Uma empresa metal-mecânica produz um tipo especial de motor. A quantidade em estoque desse motor segue uma distribuição normal com média de 200 unidades e desvio-padrão de 20. O gráfico a seguir representa a distribuição normal padrão (média igual a 0 e desvio-padrão igual a 1), em que as porcentagens representam as probabilidades entre os valores de desvio-padrão.

Qual é a probabilidade de, em um dado momento, o estoque da empresa apresentar mais de 220 unidades?



A) 84,13%.

B) 68,26%.

C) 34,13%.

D) 15,87%.

E) 13,60%.

Resolução desta questão na Plataforma.

FIGURAS E ILUSTRAÇÕES

Gráfico 1

Colunas para as frequências referentes à variável “sabor”.

Gráfico 2

Colunas para as frequências referentes à variável “tamanho”.

Gráfico 3

Colunas para as frequências referentes à variável “preço”.

Gráfico 4

Colunas para as frequências referentes à variável “vendas”.

Gráfico 5

Diagrama circular para a variável “sabor”.

Gráfico 6

Diagrama circular para a variável “tamanho”.

Gráfico 7

Diagrama circular para a variável “preço”.

Gráfico 8

Diagrama circular para a variável “vendas”.

Gráfico 9

Representação de dois conjuntos com interseção.

Gráfico 10

Representação gráfica de dois conjuntos com interseção pertencentes a um conjunto universo com elementos que não pertencem a A ou a B.

Gráfico 11

Representação de dois conjuntos sem interseção e conjunto universo indicado explicitamente.

Gráfico 12

Representação de um conjunto contido em outro.

Gráfico 13

Representação de três conjuntos sem interseção.

Gráfico 14

Representação com a contagem de elementos dos conjuntos dados no gráfico 9.

Gráfico 15

Representação com a contagem de elementos dos conjuntos dados no gráfico 10.

Gráfico 16

Representação com a contagem de elementos dos conjuntos dados no gráfico 11.

Gráfico 17

Representação com a contagem de elementos dos conjuntos dados no gráfico 12.

Gráfico 18

Representação com a contagem de elementos dos conjuntos dados no gráfico 13.

Gráfico 19

Representação com a contagem de elementos dos conjuntos dados no gráfico 13, enfatizando o conjunto A.

Gráfico 20

Representação com a contagem de elementos dos conjuntos dados no gráfico 13, enfatizando os elementos que pertencem a cada um dos conjuntos A, B e C.

Gráfico 21

Representação com a contagem de elementos dos conjuntos dados no gráfico 13, enfatizando os elementos que pertencem a B e C.

Gráfico 22

Representação com a contagem de elementos dos conjuntos dados no gráfico 13, enfatizando os elementos que não pertencem a B.

Tabela 1

Dados brutos para a compra de sucos.

Tabela 2

Frequências e frequências relativas para a variável "sabor".

Tabela 3

Frequências e frequências relativas para a variável "tamanho".

Tabela 4

Frequências sem perda de informação para a variável "preço".

Tabela 5

Frequências para a variável "preço", com agrupamento por intervalos.

Tabela 6

Frequências sem perda de informação para a variável "vendas".

Tabela 7

Frequências para a variável "vendas", com agrupamento por intervalos de tamanho 5.

Tabela 8

Frequências para a variável "vendas", com agrupamento por intervalos de tamanho 3.

Tabela 9

Quantidade de impressoras e multifuncionais vendida na loja A.

Tabela 10

Quantidade de impressoras e multifuncionais vendida na loja A.

Tabela 11

Quantidade de impressoras e multifuncionais vendida na loja A.

Tabela 12

Idades dos integrantes de uma família.

Tabela 13

Gastos efetuados para a realização do jantar por participante do grupo de amigos.

Tabela 14

Gastos efetuados por ocupante da mesa da lanchonete.

Tabela 15

Frequência para os gastos efetuados por ocupante da mesa da lanchonete.

Tabela 16

Quadro esquemático para cálculo de médias a partir de tabela de frequências.

Tabela 17

Frequências para os valores de uma variável x genérica.

Tabela 18

Cálculo do valor médio de um conjunto de dados para uma variável x genérica a partir de sua tabela de frequências.

Tabela 19

Cálculo do valor médio de um conjunto de dados a partir da tabela de frequências dada por intervalos de valores.

Tabela 20

Exemplo de cálculo de médias com ponderação: notas de alunos.

Tabela 21

Exemplo de cálculo de média aritmética para um conjunto de aumento de preços (em valores percentuais).

REFERÊNCIAS

Textuais

CRESPO, A. A. *Estatística fácil*. 18. ed., São Paulo: Editora Saraiva, 2002.

DOWNING, D.; CLARK, J. *Estatística aplicada*. São Paulo: Editora Saraiva, 1999.

FONSECA, J. S.; MARTINS G. A.; TOLEDO, G. L. *Estatística aplicada*. São Paulo: Atlas, 1995.

GUERRA, M. J.; DONAIRE, D. *Estatística aplicada.* São Paulo: Ciência e Tecnologia, 1991.

KUME, H. *Métodos estatísticos para a melhoria da qualidade.* São Paulo: Gente, 1993.

MEDEIROS, E. *et al. Estatística para os cursos de economia, administração e ciências contábeis*, 3. ed. São Paulo: Atlas, 1995/1999.

MILONE, G.; ANGELINI, F. *Estatística aplicada*. São Paulo: Atlas, 1995.

SILVA, E. M. *et al. Tabelas de estatística*. 2. ed. São Paulo: Atlas, 1999.

STEVENSON, W. J. *Estatística aplicada à administração*, São Paulo: Harbra, 1983.

VIEIRA, S. *Elementos de estatística*. 4. ed. São Paulo: Atlas, 2003.

Exercícios

Unidade VI – Questão 2: INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). *Exame Nacional de Desempenho dos Estudantes (ENADE) 2009*: Administração. Questão 32. Disponível em: < http://public.inep.gov.br/enade2009/ADMINISTRACAO.pdf>. Acesso em: 06 jun. 2011.

Informações:
www.sepi.unip.br ou 0800 010 9000