

AULA 06

Probabilidade

Ernesto F. L. Amaral

26 de agosto de 2010
Metodologia de Pesquisa (DCP 854B)

Fonte:

Triola, Mario F. 2008. “Introdução à estatística”. 10^a ed. Rio de Janeiro: LTC. Capítulo 4 (pp.110-157).

ESQUEMA DA AULA

- Fundamentos
- Regra da adição
- Regra da multiplicação: idéias básicas
- Regra da multiplicação: complementares e probabilidade condicional
- Probabilidades através de simulações
- Contagem

PROBABILIDADE

- A probabilidade é a base sobre a qual são construídos importantes métodos de inferência estatística.
- Regra do evento raro para inferência estatística: se, sob uma dada hipótese, a probabilidade de um evento particular observado for muito pequena, concluimos que, provavelmente, a hipótese não é correta.
- Objetivo principal é de entender valores de probabilidade, os quais serão úteis nos capítulos seguintes.
- Também aprenderemos como determinar valores de probabilidades em uma variedade de circunstâncias.

FUNDAMENTOS

DEFINIÇÕES

- Os valores de probabilidade se expressam como números entre 0 e 1 (inclusive).
- O importante é aprender a interpretar valores de probabilidade.
- Uma probabilidade muito pequena (0,001, por exemplo) indica que determinado evento raramente ocorre.
- Um **evento** é qualquer conjunto de resultados ou consequências de um experimento.
- Um **evento simples** é um resultado ou um evento que não pode mais ser decomposto em componentes mais simples.
- O **espaço amostral** de um experimento consiste em todos os eventos simples possíveis, ou seja, são todos resultados que não podem mais ser decompostos.

EXEMPLOS

Procedimento	Exemplo de evento	Espaço amostral
1 nascimento	<p>Evento simples: sexo feminino</p>	<p>Com 1 nascimento, há 2 resultados que são eventos simples:</p> <p>{f, m}</p>
3 nascimentos	<p>Evento: 2 femininos e 1 masculino</p> <p>Eventos simples (todos eventos simples resultantes de 2 femininos e 1 masculino): ffm, fmf, mff</p>	<p>Com 3 nascimentos, há 8 resultados que são eventos simples:</p> <p>{fff, ffm, fmf, fmm, mff, mfm, mmf, mmm}</p>

NOTAÇÃO BÁSICA PARA PROBABILIDADE

- P representa a probabilidade.
- A , B e C representam eventos específicos.
- $P(A)$ representa a probabilidade de ocorrência do evento A .

REGRAS PARA DEFINIR PROBABILIDADE DE EVENTO

- Há diferentes formas de definir a probabilidade de um evento, tais como:

REGRA 1

Aproximação da probabilidade pela frequência relativa:

$$P(A) = \frac{(n^{\circ} \text{ vezes em que ocorreu } A)}{(n^{\circ} \text{ vezes que procedimento foi repetido})}$$

- **Lei dos grandes números:** à medida que um experimento é repetido várias vezes (maior amostra), essa probabilidade tende a se aproximar da verdadeira probabilidade.
- Exemplo é a probabilidade de ocorrências de cara, ao lançar uma moeda.

REGRA 2

Abordagem clássica da probabilidade:

- Determinado experimento tem n diferentes eventos simples e cada um desses eventos simples tem igual chance de ocorrer (resultados igualmente prováveis).
- Se evento A pode ocorrer em s dessas n maneiras, então:

$$P(A) = \frac{(n^{\circ} \text{ maneiras em que } A \text{ pode ocorrer})}{(n^{\circ} \text{ diferentes eventos simples})} = s / n$$

- Exemplo é o número de maneiras em que 4 pode ocorrer ao lançar dois dados.

REGRA 3

Probabilidades subjetivas:

- $P(A)$ é estimada com base no conhecimento de circunstâncias relevantes.
- Exemplo é a previsão meteorológica para o dia seguinte.

CONSIDERAÇÕES IMPORTANTES

- Um **erro comum** consiste em, incorretamente, admitir que os resultados são igualmente prováveis porque não sabemos coisa alguma sobre a probabilidade de cada resultado.
- Quando não se sabe coisa alguma a respeito da probabilidade de diferentes resultados possíveis, **não se deve supor** que sejam igualmente prováveis.
- Nos problemas de probabilidade básica, é muito importante **examinar a informação disponível** cuidadosamente e identificar o número total de resultados possíveis.
- A **precisão dos resultados** depende da qualidade do método de amostragem e dos procedimentos de pesquisa.
- **Simulação** do experimento é um processo que se comporta da mesma maneira que o experimento, com resultados semelhantes e mais fáceis de calcular.

VALORES POSSÍVEIS DE PROBABILIDADE

- A probabilidade matemática de qualquer evento é 0, 1 ou um número entre 0 e 1:
 - A probabilidade de um evento impossível é 0.
 - A probabilidade de um evento cuja ocorrência é certa é 1.
 - Para qualquer evento A, a probabilidade de A está entre 0 e 1, inclusive ($0 \leq P(A) \leq 1$).
- Expressões mais familiares e comuns de verossimilhança:
 - Impossível: $P(A)=0$
 - Improvável: $P(A) \sim 0,25$
 - Chance 50-50: $P(A)=0,5$
 - Provável: $P(A) \sim 0,75$
 - Certo: $P(A)=1$

EVENTOS COMPLEMENTARES

- O complementar de um evento A , representado por \bar{A} , consiste em todos os resultados em que A não ocorre.

VALORES P

- Mais adiante, veremos a expressão “valor P ” com “significância inferior a 0,001” ($p < 0,001$ ou significativa a 99,9%).
- O importante é saber que uma probabilidade de 0,001 corresponde a um evento tão raro que ocorre, em média, apenas uma vez em cada mil tentativas.

ARREDONDAMENTO DE PROBABILIDADES

- Ao expressar o valor de uma probabilidade, deve-se indicar:
 - 1) A fração exata, por exemplo, $1/3$.
 - 2) O decimal exato, por exemplo, 0,5 (e não 0,500).
 - 3) Arredondar o resultado final para três algarismos significativos, sendo que todos algarismos são significativos, menos os zeros que são incluídos para o posicionamento correto da vírgula decimal (por exemplo, 0,0215, ao invés de 0,021491).

- Quando uma probabilidade não é uma fração simples ($432/7842$, por exemplo), devemos expressá-la na forma decimal (0,0551) para facilitar compreensão.

CHANCES

- As expressões de verossimilhança (probabilidade) são frequentemente dadas em forma de chances, ex.: “50:1”.
- Uma desvantagem séria das chances é que elas tornam muitos cálculos extremamente difíceis.
- A **chance real contra** a ocorrência do evento A é dada pela razão $P(\bar{A})/P(A)$, usualmente expressa na forma $a:b$ (ou “ a para b ”), onde a e b são inteiros primos entre si.
- A **chance real a favor** do evento A é o inverso da chance real contra aquele evento. Se a chance contra A é $a:b$, então a chance a favor de A é $b:a$ ou $P(A)/P(\bar{A})$.
- A **chance no rateio** contra o evento A representa a razão do lucro líquido (se você ganhar) para a quantia apostada:
(lucro líquido) : (quantia apostada)

EXEMPLO

- Digamos: (1) você aposta 5 dólares no número 13 em uma roleta; (2) sua probabilidade de ganhar é de $1/38$; e (3) a chance no rateio dada pelo cassino é de 35:1.
- **Chance real contra 13**
 - $P(13)=1/38$ e $P(\text{não } 13)=37/38$
 - $P(\text{não } 13)/P(13) = (37/38)/(1/38) = 37/1 = 37:1$
- **Lucro líquido**
 - $35:1 = (\text{lucro líquido}):(\text{quantia apostada})$
 - Lucro líquido é de \$35 para cada dólar apostado.
 - Se aposta é de \$5, apostador recebe \$180 $[(5*35)+5]$.
- **Chance no rateio = chance real contra 13**
 - Lucro líquido seria de \$37 para cada dólar apostado.
 - Cassino está lucrando \$2 para cada dólar apostado.

REGRA DA ADIÇÃO

REGRA DA ADIÇÃO (ou)

- A regra da adição é uma ferramenta para achar probabilidades que podem ser expressas como $P(A \text{ ou } B)$:
 - A probabilidade de que ocorra: (1) o evento A ; (2) o evento B ; ou (3) ambos ocorram.
- Precisamos encontrar o número total de maneiras que o evento A pode ocorrer e que o evento B pode ocorrer, mas sem contar qualquer resultado mais de uma vez.
- Usaremos mais o “**ou inclusivo**” (ou um, ou outro, ou ambos), ao invés do “**ou exclusivo**” (ou um, ou outro, mas não ambos).
- É importante saber que **evento composto** é qualquer evento que combina dois ou mais eventos simples.

$$P(A \text{ ou } B) = P(\text{evento } A, \text{ ou } B, \text{ ou ambos em única prova})$$

EXEMPLO

Resultado do teste	Sujeito realmente usou maconha?	
	Sim	Não
Positivo	119 (positivo verdadeiro)	24 (falso positivo)
Negativo	3 (falso negativo)	154 (negativo verdadeiro)

- **Qual a probabilidade de ser selecionado um sujeito que teve teste positivo ou usava maconha?**
- Somente positivo (24), somente maconha (3), ambos (119).
- $P(\text{teste positivo ou usava maconha}) = 146 / 300 = 0,487$.
- Tomou-se o cuidado de não realizar contagens duplas.

REGRA FORMAL E REGRA INTUITIVA

– Regra formal da adição:

$$P(A \text{ ou } B) = P(A) + P(B) - P(A \text{ e } B)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- $P(A \text{ e } B)$ representa a probabilidade de A e B ocorrerem em conjunto, como resultado de 1 prova do experimento.

– Regra intuitiva da adição:

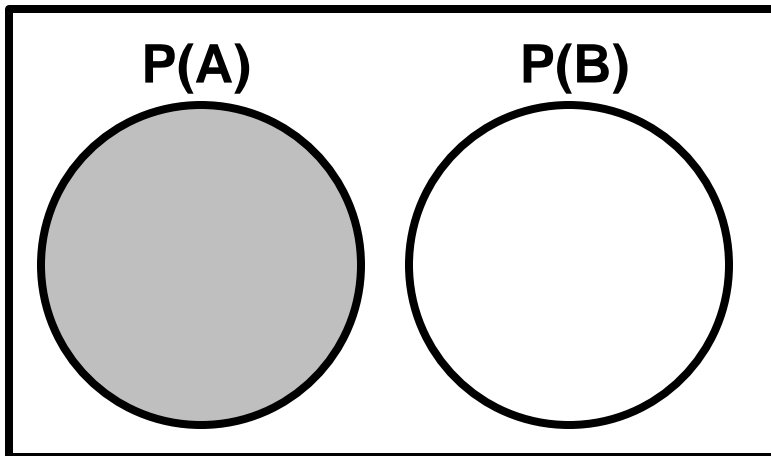
- Para achar $P(A \text{ ou } B)$, ache a soma do número de maneiras segundo as quais o evento A pode ocorrer e o número de maneiras segundo as quais o evento B pode ocorrer, somando de tal maneira que cada resultado seja contado apenas uma vez.
- $P(A \text{ ou } B)$ é igual a esta soma dividida pelo número total de resultados do espaço amostral.

DIAGRAMA DE VENN

- Eventos A e B são disjuntos (ou mutuamente exclusivos) se eles não podem ocorrer simultaneamente.
- Ou seja, eventos disjuntos não se superpõem.

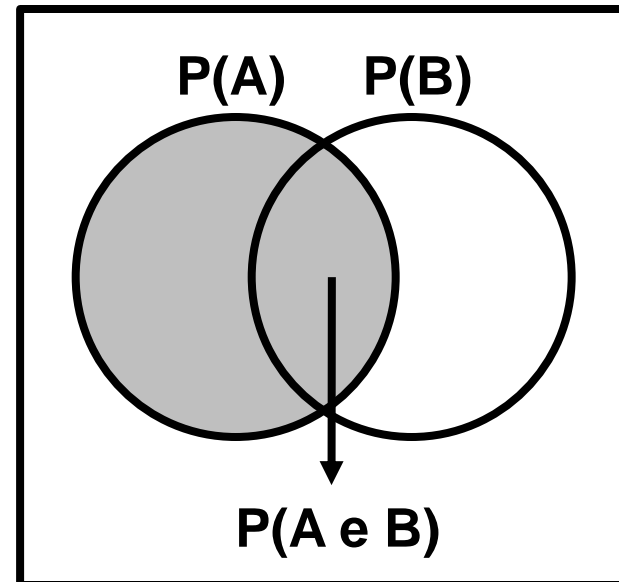
EVENTOS DISJUNTOS

Área Total = 1



EVENTOS NÃO-DISJUNTOS

Área Total = 1



$P(A \text{ e } B)$: probabilidade de que A e B ocorram ambos na mesma prova do experimento.

EVENTOS COMPLEMENTARES

- O evento A e seu complementar (\bar{A}) têm que ser disjuntos, porque é impossível um evento e seu complementar ocorrerem ao mesmo tempo.
- Podemos afirmar que A ocorre ou não ocorre, o que implica que ou A ou \bar{A} tem que ocorrer.
- Regra da adição para eventos disjuntos:

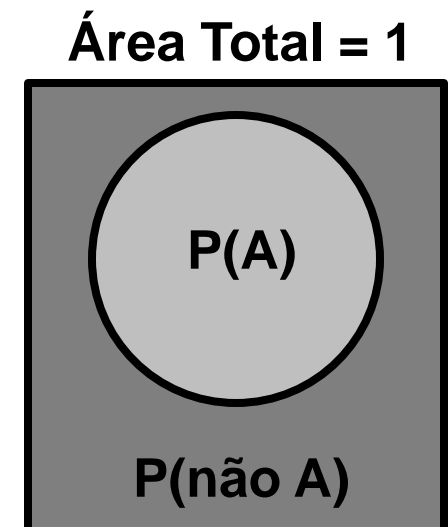
$$P(A \text{ ou } \bar{A}) = P(A) + P(\bar{A}) = 1$$

- Três expressões equivalentes:

$$P(A) + P(\bar{A}) = 1$$

$$P(\bar{A}) = 1 - P(A)$$

$$P(A) = 1 - P(\bar{A})$$



REGRA DA MULTIPLICAÇÃO: IDÉIAS BÁSICAS

REGRA DA MULTIPLICAÇÃO (e): IDÉIAS BÁSICAS

- A regra básica da multiplicação é usada para se encontrar $P(A \text{ e } B)$, a probabilidade de o evento A acontecer em uma primeira prova e o evento B ocorrer em uma segunda prova.
- Se o resultado do primeiro evento A afeta a probabilidade do segundo evento B , é importante ajustar a probabilidade de B para refletir a ocorrência do evento A .
- **Probabilidade condicional:** $P(B|A)$ representa a probabilidade do evento B ocorrer depois que se admite que o evento A ocorreu.

EXEMPLO

Resultado do teste	Sujeito realmente usou maconha?	
	Sim	Não
Positivo	119 (positivo verdadeiro)	24 (falso positivo)
Negativo	3 (falso negativo)	154 (negativo verdadeiro)

– Qual a probabilidade de que a primeira pessoa selecionada tenha um resultado de teste positivo e a segunda pessoa tenha um teste negativo?

1) $P(\text{teste positivo}) = 143/300$.

2) $P(\text{teste negativo}) = 157/299$.

– $P(1^\circ \text{ positivo e } 2^\circ \text{ negativo}) = (143/300) \times (157/299) = 0,250$.

DEFINIÇÕES IMPORTANTES

- Dois eventos A e B são **independentes** se a ocorrência de um não afeta ocorrência do outro (com reposição):

$$P(A \text{ e } B) = P(A) \times P(B)$$

- Se a ocorrência de B depende da ocorrência de A , estes eventos são **dependentes** (sem reposição):

$$P(A \text{ e } B) = P(A) \times P(B|A)$$

- **Regra intuitiva da multiplicação:**

- Ao calcular a probabilidade de ocorrência do evento A em uma prova e do evento B na prova seguinte:
 - Multiplique a probabilidade do evento A pela probabilidade do evento B .
 - Mas certifique-se de que a probabilidade do evento B leva em conta a ocorrência prévia do evento A .

VÁRIOS EVENTOS

- A probabilidade de qualquer sequência de eventos independentes é o produto das probabilidades correspondentes.
- Podemos também estender a regra da multiplicação de modo que ela se aplique a eventos dependentes, ajustando as probabilidades à medida que avançamos.

EVENTOS DEPENDENTES COMO INDEPENDENTES

- É prática comum considerarem-se os eventos como independentes quando pequenas amostras são retiradas de grandes populações:
 - É raro selecionar o mesmo item duas vezes.
- Se o tamanho da amostra não é maior que 5% do tamanho da população, trate as seleções como sendo independentes:
 - Isso é realizado mesmo que as seleções sejam feitas sem reposição, ou seja, sejam tecnicamente dependentes.
- Isso é usado em pesquisas de opinião pública, quando há poucas entrevistas em uma população de milhões:
 - Mesmo sem reposição, é considerada independência.

REGRA DA MULTIPLICAÇÃO: COMPLEMENTARES E PROBABILIDADE CONDICIONAL

REGRA DA MULTIPLICAÇÃO: COMPLEMENTARES E PROBABILIDADE CONDICIONAL

– Probabilidade complementar:

- Quando desejamos achar a probabilidade de que, entre várias tentativas, obtemos pelo menos um de alguns eventos especificados:
 - Podemos achar a probabilidade de que nenhum daqueles eventos ocorrerá.
 - Então achamos a probabilidade complementar.

– Probabilidade condicional:

- É a probabilidade de um evento, dada a informação adicional de que algum outro evento já ocorreu.

COMPLEMENTARES: PROBABILIDADE DE “PELO MENOS UM”

- A regra da multiplicação e a regra do complementar podem ser usadas em conjunto para resolver certos problemas.
- Ache a probabilidade de que, entre várias tentativas, **pelo menos um** (um ou mais) forneça um resultado especificado.
- O complementar de se obter pelo menos um de um item particular é não se obter qualquer item daquele tipo.
- Probabilidade de pelo menos um de alguma coisa é a diferença entre 1 e a probabilidade de nenhum:

$$P(\text{pelo menos um}) = 1 - P(\text{nenhum})$$

EXEMPLO: SEXO DE CRIANÇAS

– Sendo meninos e meninas igualmente prováveis e sexo de uma criança independente do sexo de outra, qual é a probabilidade de pelo menos 1 menina em 3 crianças?

1) $P(A)$ = pelo menos 1 menina em 3 crianças

2.1) $P(\text{não } A)$ = não se obter pelo menos 1 menina em 3

2.2) $P(\text{não } A)$ = todas 3 crianças são meninos

2.3) $P(\text{não } A)$ = menino e menino e menino

3.1) Probabilidade complementar = $P(\text{não } A)$

3.2) $P(\text{menino, menino, menino}) = 1/2 \times 1/2 \times 1/2 = 1/8$

4) $P(A) = 1 - P(\text{não } A) = 1 - 1/8 = 7/8$

PROBABILIDADE CONDICIONAL

- A probabilidade condicional de um evento é usada quando a probabilidade é afetada pelo conhecimento de outras circunstâncias.
- Ou seja, é a probabilidade obtida com a informação adicional de que algum outro evento já ocorreu.
- $P(B|A)$ representa a probabilidade condicional da ocorrência do evento B , dado que o evento A já ocorreu:
 - Como: $P(A \text{ e } B) = P(A) \times P(B|A)$
 - Temos: $P(B|A) = P(A \text{ e } B) / P(A)$
- **Abordagem intuitiva:** a probabilidade condicional de B dado A pode ser calculada considerando-se que o evento A ocorreu e calcular a probabilidade de que o evento B ocorrerá.

EXEMPLO

Resultado do teste	Sujeito realmente usou maconha?	
	Sim	Não
Positivo	119 (positivo verdadeiro)	24 (falso positivo)
Negativo	3 (falso negativo)	154 (negativo verdadeiro)

– **Ao escolher 1 pessoa, qual a probabilidade do teste ser positivo, visto que esta pessoa usou maconha?**

$$1) P(\text{positivo}|\text{maconha}) = 119/122 = 0,975.$$

$$= P(\text{positivo e maconha})/P(\text{maconha}) = (119/300) / (122/300)$$

$$2) P(\text{maconha}|\text{positivo}) = 119/143 = 0,832$$

$$= P(\text{maconha e positivo})/P(\text{positivo}) = (119/300) / (143/300)$$

PROBABILIDADES ATRAVÉS DE SIMULAÇÕES

PROBABILIDADES ATRAVÉS DE SIMULAÇÕES

- Ao invés de usarmos regras formais para encontrar probabilidades, podemos desenvolver uma simulação, a qual se comporta da mesma maneira que o procedimento em análise.
- Uma **simulação** de um experimento é um processo que tem o mesmo comportamento do experimento, de modo que são gerados resultados semelhantes.
- É extremamente importante que a elaboração de uma simulação seja feita de modo que ela se comporte exatamente igual ao experimento real.
- Isso pode ser feito com tabela de números aleatórios ou com programas estatísticos:
 - P(data de nascimento igual: 1 a 365) com $n=25$ no Excel.

CONTAGEM

CONTAGEM

- Em muitos problemas de probabilidade, a maior dificuldade é encontrar o número total de resultados.
- Há diferentes métodos para se encontrar tais números.
- **Princípio fundamental da contagem:** para uma sequência de dois eventos, na qual o primeiro evento pode ocorrer de m maneiras e o segundo pode ocorrer de n maneiras, os eventos juntos podem ocorrer em um total de $m*n$ maneiras.
- Exemplo:
 - Probabilidade de gerar um número aleatório de CPF.
 - 11 dígitos, sendo que cada um tem 10 resultados possíveis (0 a 9).
 - $10*10*10*10*10*10*10*10*10*10*10 = 100.000.000.000$
 - $P(\text{n}^\circ \text{ aleatório CPF}) = 1/100.000.000.000$

REGRA DO FATORIAL

- O símbolo fatorial (!) representa o produto de inteiros positivos decrescentes ($4! = 4*3*2*1 = 24$).
 - Por definição, $0! = 1$.
- Um conjunto de n diferentes itens pode ser organizado em ordem de $n!$ maneiras diferentes.
 - Isso ocorre porque o primeiro item pode ser selecionado de n diferentes maneiras, o segundo de $n-1$ maneiras...
- Exemplo:
 - Se temos que realizar pesquisas nas capitais estaduais, qual o número de diferentes rotas possíveis?
 - $27! = 27*26*25*24*23*...*3*2*1$
 - 10.888.869.450.418.400.000.000.000.000 rotas possíveis

REGRA DAS PERMUTAÇÕES (QUANDO TODOS ITENS SÃO DIFERENTES)

- Na permutação (arranjo, sequência), a ordem é levada em conta, no sentido de que diferentes ordenações dos mesmos itens são contadas separadamente.
- Requisitos:
 - Há um total de n diferentes itens disponíveis.
 - Selecionamos r dos n itens (sem reposição).
 - Temos que considerar reorganizações dos mesmos itens como sendo sequências diferentes (ABC≠ACB≠CBA...).
- Número de **permutações** (ou sequências) de r itens selecionados (sem reposição) dentre os n diferentes itens disponíveis é:

$$P_r^n = \frac{n!}{(n - r)!}$$

EXEMPLO DE PERMUTAÇÃO (QUANDO TODOS ITENS SÃO DIFERENTES)

- Se temos que realizar pesquisas nas capitais estaduais, mas dispomos de tempo para visitar apenas quatro capitais, qual o número de diferentes rotas possíveis?
- Sendo $n=27$ e $r=4$, aplicamos a fórmula:
 - $n! / (n-r)! =$
 - $27! / (27-4)! =$
 - $27! / 23! =$
 - $27 * 26 * 25 * 24 * 23! / 23! =$
 - $27 * 26 * 25 * 24 =$
 - 421.200 rotas possíveis

REGRA DAS PERMUTAÇÕES (QUANDO ALGUNS ITENS SÃO IGUAIS A OUTROS)

- Requisitos:
 - Há n itens disponíveis e alguns itens são iguais a outros.
 - Seleccionamos todos os n itens (sem reposição).
 - Consideramos os rearranjos de itens distintos como sequências diferentes.
- Se os requisitos são satisfeitos e se há n_1 iguais entre si, n_2 iguais entre si, ..., n_k iguais entre si, o número de **permutações** (ou sequências) de todos os n itens seleccionados sem reposição é:
$$\frac{n!}{n_1! n_2! \dots n_k!}$$
- Quando há apenas duas categorias, podemos definir que x sejam iguais entre si e os outros $n-x$ também:
$$\frac{n!}{(n-x)! x!}$$

EXEMPLO DE PERMUTAÇÃO (QUANDO ALGUNS ITENS SÃO IGUAIS A OUTROS)

- Um pesquisador realiza um teste de um método de seleção de sexo com 10 casais. Os resultados dos nascimentos são de 8 meninas e 2 meninos.
- Quantas maneiras diferentes 8 meninas e 2 meninos podem ser arranjados em sequência?
 - Temos $n=10$ nascimentos.
 - n_1 iguais (meninas) = 8
 - n_2 iguais (meninos) = 2
 - $n! / (n_1! n_2!) = 10! / (8! 2!) = 10 \cdot 9 \cdot 8! / (8! 2!) = 10 \cdot 9 / 2 = 45$
 - $n! / [(n-x)! x!] = 10! / [(10-8)! 8!] = 10! / (2! 8!) = 45$
 - Há 45 maneiras diferentes em que 8 meninas e 2 meninos podem ser arranjados.

PERMUTAÇÃO ≠ COMBINAÇÃO

- Quando diferentes ordenações dos mesmos itens são contadas separadamente, tem-se um problema de **permutação**:
 - Consideramos reorganizações dos mesmos itens como sendo sequências diferentes ($ABC \neq ACB \neq CBA \dots$).
- Quando as diferentes ordenações dos mesmos itens não são contadas separadamente, tem-se um problema de **combinação**:
 - Consideramos reorganizações dos mesmos itens como sendo sequências iguais ($ABC = ACB = CBA \dots$).

REGRA DAS COMBINAÇÕES

- Requisitos:
 - Há n diferentes itens disponíveis.
 - Seleccionamos r dos n itens (sem reposição).
 - Consideramos reorganizações dos mesmos itens como sendo a mesma (ABC=ACB=CBA...).
- Se os requisitos precedentes forem satisfeitos, o número de **combinações** de r itens escolhidos dentre n itens diferentes é:

$$C_r^n = \frac{n!}{(n-r)!r!}$$

EXEMPLO

– Desejamos tratar 8 pessoas sadias (r) com uma nova droga e temos 10 voluntários (n).

– **8 sujeitos são selecionados dentre 10 e tratados em sequência. Se houver reação adversa, teste é interrompido. Quantos arranjos possíveis?**

Ordem importa (rearranjos de mesmos itens são diferentes):

$$\text{permutação: } n!/(n-r)! = 10!/(10-8)! = 10!/2! = 1.814.400$$

– **8 sujeitos são selecionados dentre 10 e tratados ao mesmo tempo. Quantos arranjos possíveis?**

Ordem não importa (rearranjos de mesmos itens são iguais):

$$\text{combinação: } n!/(n-r)!r! = 10!/(10-8)!8! = 10!/2!8! = 45$$