**Estatística e Probabilidade Descritiva**

Uma estatística descritiva é um resumo de estatística que descreve quantitativamente ou resume características de uma coleção de informação. Estatísticas descritivas são apenas descritivas. Elas não procuram generalizar além dos dados disponíveis. Estatísticas descritivas providenciam simples resumos sobre a amostra e sobre as observações que foram feitas.

1. **O que é probabilidade?**

Descrevendo em termos básicos, é um conceito intuitivo. Nós usamos diariamente sem necessariamente perceber que estamos falando e aplicando probabilidade para o trabalho.

A vida é cheia de incertezas. Nós não sabemos o resultado de uma situação até que ela aconteça. Vai chover hoje? Vou passar no próximo teste? Meu time vai ganhar o jogo? Eu vou ganhar uma promoção nos próximos 6 meses? Todas essas questões são exemplos de situações incertas que vivemos. Vamos mapeá-las para uma terminologia comum que usaremos daqui para frente.

* **Experiência –** São as situações incertas, onde tem múltiplos resultados. Chover diariamente é uma experiência.
* **Resultado** é o resultado de um teste único. Então, se hoje chover, o resultado de hoje do experimento será “Choveu”.
* **Evento** é um ou mais de um resultado de um experimento. “Choveu” é um dos possíveis eventos para esse experimento.
* **Probabilidade** é uma medida da probabilidade de um evento. Então, se existe 60% de chance que vai chover amanhã, a probabilidade do resultado “Choveu” para amanhã é 0.6.

**Por que precisamos de probabilidade?**

Em um mundo incerto, pode ser uma imensa ajuda saber e conhecer chances de vários eventos. Você pode planejar coisas adequadamente. Se é provável que vai chover, eu poderia levar o meu guarda-chuva. Se é provável que eu tenha diabetes de acordo com a minha alimentação, eu vou fazer uma consulta. Se meu cliente não puder me pagar um prêmio de renovação sem um lembrete, eu vou lembrar ele disso.

Então, conhecer a probabilidade pode ser de grande benefício.

1. **Variáveis aleatórias**

Para calcular a probabilidade da ocorrência de um evento, nos precisamos colocar uma estrutura para expressar o resultado em números. Nós conseguimos fazer isso mapeando o resultado de um experimento em números.

Vamos definir X para ser o resultado de um lance de moeda.

X = resultado de um lance de moeda.

Possíveis resultados:

* 1 se sair “cara”
* 0 se sair “coroa”

Vamos ver outro exemplo.

Supomos que, eu ganho o jogo se tiver uma soma de 8 enquanto rolo dois dados iguais. Eu posso definir minha variável aleatória Y para ser (a soma do lado de cima dos dois dados).

Y pode ter os valores = (2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)

Algumas coisas para anotar sobre variáveis aleatórias:

* Cada valor de uma variável aleatória pode ou não pode ser igualmente provável. Há somente 1 combinação de dado, onde a soma dá 2[(1,1)], enquanto uma soma de 5 pode ser resultado de [(1,4), (2,3), (3,2), (4,1)]. Então, 5 é mais provável de ocorrer do que comparado com 2. Pelo contrário, a probabilidade de cara ou coroa no jogar de uma moeda é igualmente 50-50.
* Algumas vezes, as variáveis aleatórias podem ter apenas valores fixos, ou valores somente em um certo intervalo. Por exemplo em um dado, o lado de cima vai somente mostrar valores entre 1 e 6. Não pode ter um valor como 2.25 ou 1.5. Similarmente, quando uma moeda é jogada, somente será mostrado cara ou coroa e nada mais. Por outro lado, se eu definir minha variável aleatória para ser a quantidade de açúcar em uma laranja, pode ter qualquer valor como 1.4g, 1.45g, 1.456g, 1.4568g e por aí vai. Todos esses valores são possíveis e todos os infinitos valores entre eles também são possíveis. Então, nesse caso, a variável aleatória é contínua com a possibilidade de todos os números reais.
* Não pense em variável aleatória como uma variável tradicional (mesmo que ambas sejam chamadas de variáveis) como y=x+2, onde o valor de y está dependente de x. Variável aleatória é definida em termos do resultado de um processo. Nós quantificamos o processo usando variável aleatória.

1. **Calculando probabilidade pelo princípio da contagem**

Vamos dizer que você foi para uma feira. Tem uma tenda jogando o jogo da roda giratória, onde tem duas cores uniformemente espalhado pela roda, vermelho e verde. Se você parar no vermelho, você perde, se você parar no verde você ganha.

Então o que acontece quando você gira a roda? Você ganha ou você perde? Não tem terceiro resultado nesse caso. Se a roda for justa, tem 50% de chance de ganhar e 50% de chance de perder.

Agora, supomos que o organizador decide aumentar o preço em dinheiro e reduzir a área verde. Agora somente ¼ da área é verde e ¾ é vermelho.

O quão próximo você está de vencer agora?

Somente 25%! Esse 25% ou 0.25 é a probabilidade de vencer.

**Duas jogadas de um dado**

Na próxima tenda é a nossa tenda favorita de dados, onde nós ganhamos se conseguirmos uma soma de 8 em duas jogadas. Vamos ver se nós temos mais chances de ganhar aqui.

Vamos pegar a variável aleatória X para ser a soma de duas jogadas, X pode ter valores de (2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12). Vamos ver a probabilidade de cada número.

Tem 6 possibilidades na primeira jogada (nós podemos pegar qualquer número) e 6 também na segunda jogada. Então o número total de combinações pode ser 36.

Vamos ver como:

2{(1,1)} => 1/36

3{(1,2),(2,1)} => 2/36

4{(2,2),(3,1),(1,3)} => 3/36

5{(1,4),(4,1),(2,3),(3,2)} => 4/36

6{(3,3),(1,5),(5,1),(2,4),(4,2)} => 5/36

7{(1,6),(6,1),(2,5),(5,2),(3,4),(4,3)} => 6/36

8{(2,6),(6,2),(3,5),(5,3),(4,4)} => 5/36

9{(3,6),(6,3),(5,4),(4,5)} => 4/36

10{(4,6),(6,4),(5,5)} => 3/36

11{(5,6),(6,5)} => 2/36

12{(6,6)} = > 1/36

Então, a chance de sucesso aqui é 5/36 ou aproximadamente 1 em 7; enquanto a chance de falha é 31/36. Então, ao menos que a tenda me volte 7 vezes o dinheiro que apostei para ganhar, é um jogo ruim para participar.

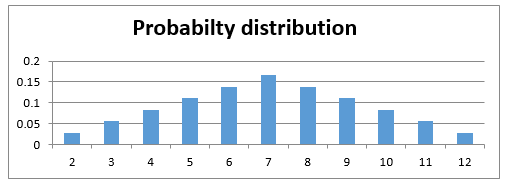
Nós podemos escrever isso como:

* P (Sucesso) = 5/36
* P (Falha) = 31/36

Você pode também ver que a probabilidade total é 1. Há somente duas possibilidades.

Vamos ver como essas probabilidades parecem. A função de probabilidade para uma variável aleatória discreta é a **função de massa de probabilidade**. Mostra as exatas probabilidades para um valor em particular da variável aleatória.

Eis uma coisa importante para anotar, a soma de 2.5 não é possível na jogada de dois dados. Então essencialmente, a minha variável aleatória é discreta. Há somente dois valores inteiros fixos que podem ter um valor e podemos ver as probabilidades de cada ocorrência.



Se você quiser aprender mais sobre probabilidade, você pode acessar [esse link](https://www.coursera.org/learn/introductiontoprobability).

Essa [folha de dicas de probabilidade](https://www.sas.upenn.edu/~astocker/lab/teaching-files/PSYC739-2016/probability_cheatsheet.pdf) te fornece um compreensivo material de referência para probabilidades e estatísticas.