Introdução

Nos capítulos 3, 4 e 5, preocupamo-nos essencialmente com o tratamento de dados amostrais, isto é, provenientes de uma amostra.

Debruçar-nos-emos um pouco, ainda que de forma por vezes meramente introdutória, sobre o estudo de populações.

Os conceitos que serão expostos ao longo deste capítulo são teóricos, isto é, provenientes de deduções matemáticas. No entanto, uma vez provado o tipo desta ou de outra população, eles mostram ser de grande utilidade e aplicabilidade prática. Além disso, ajudam-nos a compreender muitos fenómenos que ocorrem de forma natural e que, de outra forma, não poderiam ser explicados.

1. Conceito de probabilidade

1.1 A probabilidade no nosso dia-a-dia

Todos nós temos interiorizado o **conceito de probabilidade**. Por exemplo, ao sair de casa de manhã, todos nós já pensámos várias vezes na probabilidade de chover ao longo do dia, sabendo que não choverá necessariamente, mas pode chover. Também já todos associámos a um jogo de futebol a maior probabilidade de o nosso clube cor-de-rosa vencer, com a derrota do clube lilás, conscientes no entanto de que não podemos confiar neste resultado inteiramente. Até já associámos percentagens tentando quantificar o quanto podemos confiar na vitória do nosso clube. E em jogos de sorte ou azar já apostámos mediante a probabilidade de vencer, tendo consciência de que estamos a arriscar, pois apesar de termos grandes chances, podemos perder.

Esta maior ou menor confiança que depositámos num resultado futuro é o nosso conceito de **probabilidade**.

A **teoria das probabilidades**, que define e estuda a probabilidade de cada resultado, tenta quantificar a noção de *provável*.

Antes de darmos uma definição matemática (isto é mais científica) de probabilidade, convém adquirir ainda alguns conceitos essenciais em Estatística. É o que faremos na próxima secção.

1.2 Acontecimentos em Estatística

Para definirmos uma probabilidade em qualquer situação concreta, temos que definir primeiro o acontecimento ao qual vamos associar essa mesma probabilidade.

Um acontecimento no Telejornal é algo que já aconteceu, é um facto.

De forma diferente, em Estatística, quando falamos de **acontecimento** falamos de <u>algo que pode vir a acontecer ou não</u> (ou então, eventualmente, de algo que já aconteceu, mas não necessariamente).

Os acontecimentos representam-se por letras maiúsculas, por exemplo:

A = *" Amanhã chove."*

ou:

B = *"Lanço um dado com as faces numeradas de 1 a 6 e sai um 6"*

Esperamos até amanhã para ver se A ocorreu ou não ; lançamos o dado para ver se afinal sai o 6 ou não. Mas o acontecimento pode ser definido antes da experiência realizada, antes de sabermos se se iria concretizar ou não.

Se tivermos alguns conhecimentos de meteorologia, poderemos ainda hoje associar uma probabilidade a A acontecer. Essa probabilidade denota-se por P (A).

P (A) = Probabilidade de A ocorrer

Podemos fazer o mesmo com o acontecimento B e teremos então **P (B)**.

Há acontecimentos que são possíveis e acontecimentos que são impossíveis.

Um acontecimento que não vai acontecer de certeza absoluta é um **acontecimento impossível**, por exemplo:

C *= "Lanço um dado com as faces numeradas de 1 a 6 e sai um 7"*

Um acontecimento que pode eventualmente ocorrer ou não é um **acontecimento possível**. A e B (definidos atrás) são ambos acontecimentos possíveis.

Dentro da classe dos acontecimentos possíveis, temos ainda os **acontecimentos certos**, aqueles que vão ocorrer de certeza absoluta, por exemplo:

D *= "Lanço um dado com as faces numeradas de 1 a 6 e sai um número menor que 10"*

Este acontecimento tem uma fiabilidade total.

Também podemos definir acontecimentos em Estatística mediante o emprego de variáveis aleatórias (rever o início do capítulo 3). Este procedimento tem muito interesse no estudo de várias populações e no estudo das variáveis.

Por exemplo, vamos supor que definimos as variáveis:

X *= "número de filhos de um casal escolhido ao acaso durante uma certa missa católica"*

e

Y *= " altura de um estudante da ULP seleccionado ao acaso, em metros".*

NOTE BEM: Em X e em Y, não estamos a definir acontecimentos, estamos a definir variáveis, pois não fizemos qualquer afirmação da qual se possa dizer alguma vez que ocorreu ou não, nestas definições.

Então poderemos definir o acontecimento:

E *= "Um casal seleccionado ao acaso durante uma certa missa católica não tem filhos"*

ou alternativamente:

$$\boldsymbol{X} = 0$$

E, deste modo, quando quisermos associar uma probabilidade ao acontecimento E, podemos escrever:

$$P\,(\,\boldsymbol{E}\,) \text{ ou } P\,(\,\boldsymbol{X} = 0\,)$$

Também podemos definir o acontecimento:

F *= "Um estudante da ULP seleccionado ao acaso tem mais do que 1,80 m de altura"*

ou alternativamente:

$$\boldsymbol{Y} > 1{,}80$$

Note que não temos que indicar as unidades de medida, pois elas já constavam à partida na definição da variável Y.

E, deste modo, quando quisermos associar uma probabilidade ao acontecimento F, podemos escrever:

$$P\,(\,\boldsymbol{F}\,) \text{ ou } P\,(\,\boldsymbol{Y} > 1{,}80\,)$$

Uma grande vantagem da definição de acontecimentos mediante o emprego de variáveis aleatórias é que não necessitamos de estar a definir cada acontecimento sempre que o queremos tratar ou estudar. Com uma só definição da variável X podemos definir muitos acontecimentos, como:

$\mathbf{X} = 2$; $\mathbf{X} > 0$ ou $\mathbf{X} \geq 1$; $\mathbf{X} > 4$; $\mathbf{X} \leq 3$; $1 \leq \mathbf{X} < 4$; etc…

E apenas com a definição da variável Y podemos definir acontecimentos, como:

$\mathbf{Y} = 1{,}57$; $\mathbf{Y} > 2{,}10$; $\mathbf{Y} \leq 1{,}40$; $1{,}70 \leq \mathbf{Y} \leq 1{,}80$; etc…

1.3 Definições de probabilidade

Definição clássica de probabilidade

Segundo esta definição, uma vez definido um acontecimento A, a probabilidade de ele ocorrer vem dada pelo quociente entre o número de resultados possíveis que lhe são favoráveis e o número total de resultados possíveis.

$$\boxed{P\,(\,A\,) = \frac{\text{N.}^{\text{o}} \text{ de resultados possíveis favoráveis a A}}{\text{N.}^{\text{o}} \text{ total de resultados possíveis}}}$$

Exemplo – *Definidos os acontecimentos:*

A = "Sai um 6 no lançamento de um dado não viciado"

B = "Sai um número par no lançamento de um dado não viciado"

As suas probabilidades associadas são:

$$P\,(\,A\,) = \frac{1}{6} = 0{,}1667 \qquad e \qquad P\,(\,B\,) = \frac{3}{6} = \frac{1}{2} = 0{,}5000$$

Respectivamente porque há apenas 1 resultado favorável a A (a saída de um 6) e 3 resultados favoráveis a B (a saída de 2, 4 ou 6), em relação a um total de 6 resultados possíveis (1 , 2 , 3 , 4 , 5 e 6).

NOTE BEM: Esta definição só é aplicável quando todos os resultados possíveis possuem a mesma probabilidade de ocorrer, isto é, são ***equiprováveis*** !!

Exemplo - *Em relação ao exemplo anterior, o cálculo que realizámos de probabilidades não seria válido, caso o dado fosse viciado; isto é, caso as várias faces não possuíssem a mesma probabilidade de ocorrer.*

Definição frequencista de probabilidade

Segundo este conceito, obtemos a probabilidade associada a determinado acontecimento se após um grande número de repetições da experiência aleatória subjacente (*lei dos grandes números*), calcularmos a proporção de vezes em que esse acontecimento ocorre.

Assim, a abordagem frequencista do conceito de probabilidade estabelece uma correspondência entre a probabilidade de realização de um acontecimento e a frequência relativa estabilizada observada para esse acontecimento, numa amostra de muito elevada dimensão, com n a tender para infinito.

$$P(A) = \lim_{n \to \infty} f_A = \lim_{n \to \infty} \frac{n_A}{n}$$

Na prática, não é possível realizar uma experiência um número infinito de vezes. Portanto, quanto mais vezes ela for realizada, mais fiável é o cálculo da probabilidade.

Exemplo – *Definidos os acontecimentos:*

A = "Sai um 6 no lançamento de um dado viciado"

B = "Sai um número par no lançamento de um dado viciado"

O dado foi lançado 5000 vezes e contaram-se as saídas obtidas para cada face:

x_i, *face*	*1*	*2*	*3*	*4*	*5*	*6*
n_i	*123*	*345*	*523*	*887*	*1092*	*2030*
f_i	*0,0246*	*0,0690*	*0,1046*	*0,1774*	*0,2184*	*0,4060*

Assim:

$$P(A) \approx 0{,}4060$$

$$P(B) \approx 0{,}0690 + 0{,}1774 + 0{,}4060 = 0{,}6524$$

Ou então, definindo a variável aleatória:

X = "N.º saído no lançamento de um dado viciado"

Vem:

$$P(X = 6) \approx 0{,}4060$$

$$P(X \text{ é par}) \approx 0{,}0690 + 0{,}1774 + 0{,}4060 = 0{,}6524$$

Existem ainda outras definições de probabilidade que não serão abordadas aqui.

Decorre naturalmente de todas as definições de probabilidade que, para todo e qualquer acontecimento A que seja definido, é sempre:

$$\boxed{0 \le P(A) \le 1}$$

Além disso, será fácil concluir que:

Se A é um acontecimento impossível: **P (A) = 0**

Se A é um acontecimento certo: **P (A) = 1**

Se A é um acontecimento possível, mas não certo: **0 < P (A) < 1**

Exemplo - *Seja X uma variável aleatória definida do seguinte modo:*

X = "N.º saído no lançamento de um dado viciado"

Podemos afirmar à partida, sem mais estudos, que:

P (X = 7) = 0 , porque X = 7 é um acontecimento impossível.

P (X < 21) = 1 , porque X < 21 é um acontecimento certo.

P (X > 3) tem um valor maior que 0 e menor que 1, porque se trata de um acontecimento possível, mas não certo.

1.4 Algumas propriedades relacionadas com o conceito de probabilidade

Probabilidade do acontecimento complementar

O **acontecimento complementar** de outro acontecimento compreende todos os resultados em que o segundo não acontece e representa-se por $\overline{A}$. As probabilidades destes dois acontecimentos estão relacionadas por:

$$\boxed{P(\overline{A}) = 1 - P(A)}$$

Exemplo: *Seja: A = "Sai um 6 no lançamento de um dado viciado"*

Então: $\overline{A}$ = "Não sai um 6 no lançamento de um dado viciado"

Se for P (A) = 0,4060, então P ($\overline{A}$) = 1 – 0,4060 = 0,5940

Probabilidade da reunião de dois acontecimentos quaisquer

Se A e B são dois acontecimentos quaisquer, então a probabilidade de um ou outro ou ambos ocorrerem (**reunião**) vem dada por:

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

Exemplo: *Considere ainda o exemplo dado atrás sobre um dado viciado e os valores aí fornecidos.*

Sejam: *B = "Sai um número par no lançamento de um dado viciado"*

C = "Sai um número menor que 3 no lançamento de um dado viciado"

Então $B \cup C$ = "*Sai um número par ou menor que 3 no lançamento de um dado viciado*"

e $B \cap C$ = "*Sai um número par e menor que 3 no lançamento de um dado viciado" ; ou seja X = 2.*

Já vimos que $P(B) = 0{,}6524$

É fácil calcular que $P(C) = 0{,}0246 + 0{,}0690 = 0{,}0936$

Também se calcula: $P(B \cap C) = P(X = 2) = 0{,}0690$

Logo: $P(B \cup C) = P(B) + P(C) - P(B \cap C) = 0{,}6524 + 0{,}0936 - 0{,}0690 = 0{,}6770$.

Probabilidade da intersecção de dois acontecimentos independentes

Se A e B são dois acontecimentos independentes (e só nesse caso!!), isto é, se a ocorrência de um deles não altera a probabilidade de o outro ocorrer, então a probabilidade de ocorrerem simultaneamente (intersecção) vem dada por:

$$\boxed{P(A \cap B) = P(A) \times P(B)}$$

Exemplo: *Sejam:* *A = "Sai um 6 no lançamento de um dado viciado"*

D = " Sai face no lançamento de uma moeda não viciada"

Então $A \cap D$ = "*Sai um 6 no lançamento de um dado viciado e sai face no lançamento simultâneo de uma moeda não viciada"*

É fácil calcular que $P(D) = 1/2 = 0{,}5000$

Se for $P(A) = 0{,}4060$, *então* $P(A \cap D) = 0{,}4060 \times 0{,}5000 = 0{,}2030$.

1.5 Conceito de probabilidade condicionada

Por vezes pretendemos calcular a probabilidade de um acontecimento A ocorrer, no conhecimento da já ocorrência de um outro acontecimento B (que poderá influenciar a probabilidade do primeiro ou não).

Define-se assim a probabilidade condicionada e representa-se por: **P (A / B)** como sendo **a probabilidade de A ocorrer sabendo que B ocorreu**. Alternativamente, podemos dizer que P (A / B) é **a probabilidade de A ocorrer, se B ocorrer**.

P (A / B) calcula-se do seguinte modo:

$$\boxed{P(A/B) = \frac{P(A \cap B)}{P(B)}}$$

Se A e B forem acontecimentos independentes, então, a ocorrência de um deles não altera a probabilidade de ocorrência do outro e ter-se-á que:

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

Exemplo: *Considere novamente o exemplo anterior do dado viciado e também os acontecimentos A e B definidos anteriormente:*

A = "Sai um 6 no lançamento de um dado viciado"

B = "Sai um número par no lançamento de um dado viciado"

Suponha que já está em posse do conhecimento de que o lançamento do dado forneceu um número par (B ocorreu), mas não sabe que número é esse e pretende calcular a probabilidade de que ele seja um 6 ; isto é, o que pretende calcular é:

$P(A/B) = P(A \cap B) / P(B) = 0{,}4060 / 0{,}6524 = 0{,}6223$

Note que neste caso $P(A \cap B) = P(A) \approx 0{,}4060$.

Note ainda que $P(A) \neq P(A/B)$, o que mostra que A e B não são acontecimentos independentes.

Probabilidade da interseção de dois acontecimentos dependentes

Esta definição decorre naturalmente e é dedutível a partir do conceito de probabilidade condicionada.

Se A e B são dois acontecimentos dependentes, isto é, se a ocorrência de um deles altera ou pode alterar a probabilidade de o outro ocorrer, então a probabilidade de ocorrerem simultaneamente (interseção) vem dada por:

$$\boxed{P(A \cap B) = P(A) \times P(B/A) = P(B) \times P(A/B)}$$

- FIM -