ESTATÍSTICA APLICADA À ADMINISTRAÇÃO

Thiago Marzagão[1]

[1]marzagao.1@osu.edu

DISTRIBUIÇÕES CONTÍNUAS DE PROBABILIDADE

variáveis discretas vs variáveis contínuas

- Exemplos de variáveis discretas:
- ...quantidade de alunos na sala
- ...quantas vezes uma moeda cai “cara" em 10 lançamentos
- ...soma das faces de um dado em 10 lançamentos
- Exemplos de variáveis contínuas:
- ...tempo decorrido entre dois eventos
- ...peso
- ...inflação

variáveis discretas vs variáveis contínuas

- Variável discreta: pode assumir número finito de valores.
- Variável contínua: pode assumir número infinito de valores.

distribuições contínuas de probabilidade

- Uma distribuição de probabilidade é uma função, $f(x)$, que nos diz qual a probabilidade de x.
- Exemplo: litros de gasolina vendidos diariamente num posto de combustível.
- Mínimo: 0 litros.
- Máximo: 50 mil litros.
- Podemos enumerar todas as possibilidades?
- 0 litros
- 0,9 litros
- 0,99 litros
- 0,999 litros
- 0,9999... litros
- Impossível enumerar! Espaço amostral é infinito.

distribuições contínuas de probabilidade

- $f(5000) = ?$
- Aula 1: probabilidade = maneiras diferentes em que o evento pode acontecer / total de eventos que podem acontecer
- $f(5000) = 1/\infty$
- $f(5000) = 0$
- Quando a variável é contínua, a probabilidade de qualquer evento é zero.
- Mas então o que dá p/ calcular?

distribuições contínuas de probabilidade

- Quando a variável é contínua, o que dá pra medir é a probabilidade de *intervalos*.
- $f([0-25000]) = ?$
- Podemos calcular, por exemplo, a probabilidade de o posto vender entre 0 e 25 mil litros de combustível.
- Sim, é esquisito: a probabilidade de qualquer X litros individualmente é zero mas a soma de várias dessas probabilidades é maior que zero. Não tem como explicar esse paradoxo sem cálculo integral, que não vamos ver aqui.

distribuições contínuas de probabilidade

- Como exatamente calculamos a probabilidade de um intervalo?
- R: Representamos a distribuição de probabilidades em forma de gráfico e calculamos a área do gráfico entre os pontos inicial e final do intervalo.
- A fórmula exata depende da distribuição.
- (Desenhar no quadro diferentes distribuições: uniforme, exponencial, normal.)
- Aqui só veremos a distribuição normal, que é de longe a mais importante.

distribuições contínuas de probabilidade

- O que aprendemos até agora:
- Primeira lição de hoje: quando a variável é contínua, a probabilidade de qualquer evento é zero.
- Segunda lição de hoje: quando a variável é contínua, o que dá pra medir é a probabilidade de *intervalos*.
- Terceira lição de hoje: quando a variável é contínua, a probabilidade de um intervalo corresponde à área entre os pontos inicial e final.

a distribuição normal

- A distribuição normal tem formato de "sino".
- (Desenhar no quadro.)
- É a distribuição mais importante da estatística. Caracteriza inúmeros fenômenos da natureza e da vida quotidiana.
- Exemplos: peso, altura, idade, QI, salário, notas...
- (Coletar exemplos da sala.)

a distribuição normal

$$\int_a^b f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

- μ = média
- σ = desvio-padrão
- $\pi \approx$ 3,14
- $e \approx$ 2,71

a distribuição normal *padrão*

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

- Quando $\mu = 0$ e $\sigma = 1$, chamamos a distribuição normal de distribuição normal *padrão*.
- Vamos aprender a encontrar a probabilidade de um determinado intervalo na distribuição normal *padrão*.
- Depois vamos aprender a encontrar a probabilidade de um dado intervalo em qualquer distribuição normal.

a nossa “cola"

Tabela da Distribuição Normal Padrão

P(Z<z)

z	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015

exercício 1

- Trace um gráfico da distribuição normal padrão. Rotule o eixo horizontal nos valores -3, -2, -1, 0, 1, 2 e 3. Depois use a tabela de probabilidades da distribuição normal padrão para calcular as seguintes probabilidades:
- a) $P(z \leq 1,5)$
- b) $P(z \leq 1)$
- c) $P(1 \leq z \leq 1,5)$
- d) $P(0 < z < 2,5)$
- (Ex. 10 em Anderson et al, pp. 257-258)

exercício 2

- Dado que z é uma variável aleatória normal padrão, calcule as seguintes probabilidades:
- a) $P(z \leq -1,0)$
- b) $P(z \geq -1)$
- c) $P(z \geq -1,5)$
- d) $P(-2,5 \leq z)$
- e) $P(-3 < z \leq 0)$
- (Ex. 11 em Anderson et al, pp. 258)

exercício 3

- Dado que z é uma variável aleatória normal padrão, calcule as seguintes probabilidades:
- a) $P(0 \leq z \leq 0,83)$
- b) $P(-1,57 \leq z \leq 0)$
- c) $P(z > 0,44)$
- d) $P(z \geq -0,23)$
- e) $P(z < -1,20)$
- f) $P(z \leq -0,71)$
- (Ex. 12 em Anderson et al, pp. 258)

exercício 4

- Dado que z é uma variável aleatória normal padrão, calcule as seguintes probabilidades:
- a) $P(-1,98 \leq z \leq 0,49)$
- b) $P(0,52 \leq z \leq 1,22)$
- c) $P(-1,75 \leq z \leq -1,04)$
- (Ex. 13 em Anderson et al, pp. 258)

exercício 5

- Dado que z é uma variável aleatória padrão, encontre z para cada uma das situações:
- a) A área à esquerda de z é 0,9750.
- b) A área entre 0 e z é 0,4750.
- c) A área à esquerda de z é 0,7291.
- d) A área à direita de z é 0,1314.
- e) A área à esquerda de z é 0,6700.
- f) A área à direita de z é 0,3300.
- (Ex. 14 em Anderson et al, p. 258)

exercício 6

- Dado que z é uma variável aleatória padrão, encontre z para cada uma das situações:
- a) A área à esquerda de z é 0,2119.
- b) A área entre $-z$ e z é 0,9030.
- c) A área entre $-z$ e z é 0,2052.
- d) A área à esquerda de z é 0,9948.
- e) A área à direita de z é 0,6915.
- (Ex. 15 em Anderson et al, p. 258)

exercício 7

- Dado que z é uma variável aleatória padrão, encontre z para cada uma das situações:
- a) A área à direita de z é 0,01.
- b) A área à direita de z é 0,025.
- c) A área à direita de z é 0,05.
- d) A área à direita de z é 0,10.
- (Ex. 16 em Anderson et al, p. 258)

a distribuição normal

$$\int_a^b f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

- O que fazer quando $\mu \neq 0$ ou $\sigma \neq 1$?
- R: converter os pontos inicial e final p/ uma distribuição normal *padrão*.
- $z = \dfrac{x - \mu}{\sigma}$
- Daí é só consultar a nossa "cola", exatamente da mesma forma como fizemos antes.

exercício 8

- Uma variável aleatória é normalmente distribuída com uma média de $\mu = 50$ e um desvio padrão de $\sigma = 5$.
- b) Qual é a probabilidade de a variável aleatória assumir um valor entre 45 e 55?
- c) Qual é a probabilidade de a variável aleatória assumir um valor entre 40 e 60?
- (Ex. 9 em Anderson et al, p. 257, *b-c*)

exercício 9

- Para os mutuários com boas pontuações de crédito, a dívida média de contas rotativas escalonadas é de \$ 15.015 (*Business Week*, 20 de março de 2006.) Considere que o desvio padrão é \$ 3.540 e que as quantias de débito são normalmente distribuídas.
- a) Qual é a probabilidade de que o débito de um mutuário com bom crédito seja maior do que \$ 18.000?
- b) Qual é a probabilidade de que o débito de um mutuário com bom crédito seja menor do que \$ 10.000?
- c) Qual é a probabilidade de que o débito de um mutuário com bom crédito esteja entre \$ 12.000 e \$ 18.000?
- d) Qual é a probabilidade de que o débito de um mutuário seja de não mais do que \$ 14.000?
- (Ex. 17 em Anderson et al, pp. 257-258)

exercício 9 - resolução

- Primeiro passo: "traduzir" o enunciado da questão.
- $\mu = 15015$, $\sigma = 3540$
- a) $P(x > 18000) =?$
- b) $P(x < 10000) =?$
- c) $P(12000 < x < 18000) =?$
- d) $P(x < 14000) =?$

exercício 9 - resolução (cont.)

- Primeiro passo: "traduzir" o enunciado da questão.
- $\mu = 15015$, $\sigma = 3540$
- a) $P(x > 18000) = ?$
- b) $P(x < 10000) = ?$
- c) $P(12000 < x < 18000) = ?$
- d) $P(x < 14000) = ?$
- Segundo passo: converter de x p/ z
- a) $z = \frac{x - \mu}{\sigma} = \frac{18000 - 15015}{3540} = 0,865$
- b) $z = \frac{x - \mu}{\sigma} = \frac{10000 - 15015}{3540} = -1,453$
- c) $z = \frac{x - \mu}{\sigma} = \frac{12000 - 15015}{3540} = -0,873$
- $z = \frac{x - \mu}{\sigma} = \frac{18000 - 15015}{3540} = 0,865$
- d) $z = \frac{x - \mu}{\sigma} = \frac{14000 - 15015}{3540} = -0,294$

exercício 9 - resolução (cont.)

- Segundo passo: converter de x p/ z
- a) $z = \frac{x-\mu}{\sigma} = \frac{18000-15015}{3540} = 0,865$
- b) $z = \frac{x-\mu}{\sigma} = \frac{10000-15015}{3540} = -1,453$
- c) $z = \frac{x-\mu}{\sigma} = \frac{12000-15015}{3540} = -0,873$
- $z = \frac{x-\mu}{\sigma} = \frac{18000-15015}{3540} = 0,865$
- d) $z = \frac{x-\mu}{\sigma} = \frac{14000-15015}{3540} = -0,294$
- Terceiro passo: reescrever enunciado em termos de z
- a) $P(z > 0,865) =?$
- b) $P(z < -1,453) =?$
- c) $P(-0,873 < z < 0,865) =?$
- d) $P(z < -0,294) =?$

exercício 9 - resolução (cont.)

- Terceiro passo: reescrever enunciado em termos de z
- a) $P(z > 0,865) =?$
- b) $P(z < -1,453) =?$
- c) $P(-0,873 < z < 0,865) =?$
- d) $P(z < -0,294) =?$
- Quarto (e último!) passo: usar nossa “cola" p/ encontrar as probabilidades.
- a) $P(z > 0,865) = 0,1949$
- b) $P(z < -1,453) = 0,0735$
- c) $P(-0,873 < z < 0,865) = 0,8051 - 0,1922 = 0,6129$
- d) $P(z < -0,294) = 0,3859$

exercício 10

- A média de preço das ações das empresas que compõem a S&P é de $ 30 e o desvio padrão é $ 8,20 (*Business Week*, edição especial anual, primavera de 2003). Suponha que os preços das ações se distribuam normalmente.
- a) Qual é a probabilidade de uma empresa ter um preço de, no mínimo, $ 40 para suas ações?
- b) Qual é a probabilidade ter um preço não superior a $ 20 para suas ações?
- c) Qual deve ser o preço das ações para que a empresa seja incluída entre os 10% maiores?
- (Ex. 18 em Anderson et al, p. 259)

exercício 11

- Em um artigo sobre o custo dos serviços de assistência médica, a revista *Money* relatou que uma consulta em uma sala de emergência de um hospital para tratar de algo tão simples como uma dor de garganta tem um custo médio de $ 328 (*Money*, janeiro de 2009). Considere que o custo desse tipo de consulta em uma sala de emergência de um hospital seja normalmente distribuído com um desvio padrão de $ 92. Responda às seguintes questões sobre o custo do tipo de consulta mencionado anteriormente.
- a) Qual é a probabilidade de que o custo será maior do que $ 500?
- b) Qual é a probabilidade de que o custo será menor do que $ 250?
- c) Qual é a probabilidade de que o custo estará entre $ 300 e $ 400?
- d) Se o custo para um paciente estiver entre os 8% menores encargos para esse serviço médico, qual foi custo da consulta na sala de emergência para esse paciente?
- (Ex. 19 em Anderson et al, p. 259)

exercício 12

- Em janeiro de 2003 o trabalhador norte-americano passou em média 77 horas conectado à Internet enquanto se encontrava no trabalho (*CNBC*, 15 de março de 2003). Suponha que a média populacional é 77 horas, que os tempos estejam normalmente distribuídos e que o desvio padrão seja de 20 horas.
- a) Qual é a probabilidade de um trabalhador escolhido aleatoriamente ter passado menos de 50 horas conectado à Internet?
- b) Qual porcentagem de trabalhadores passou mais de 100 horas conectados à Internet?
- c) Uma pessoa é classificada como usuário intensivo se estiver entre os 20% que fazem mais uso. Quantas horas um trabalhador deve se manter conectado à Internet para ser classificado como usuário intensivo?
- (Ex. 20 em Anderson et al, p. 259)

exercício 13

- Uma pessoa deve obter uma pontuação entre os 2% mais bem classificados da população em um teste de QI para afiliar-se à Mensa, uma sociedade internacional de pessoas com QI elevado (*US Airways Attache*, setembro de 2000). Se as pontuações de QI forem normalmente distribuídas com uma média 100 e desvio padrão igual a 15, qual pontuação uma pessoa deve obter para poder afiliar-se à Mensa?
- (Ex. 21 em Anderson et al, p. 259)

exercício 14

- O hábito de assistir à TV atingiu uma nova marca quando a Nielsen Company relatou um tempo médio diário de 8,35 horas por espectador assistindo à TV (*USA Today*, 11 de novembro de 2009). Utilize uma distribuição de probabilidade normal com um desvio padrão de 2,5 horas para responder às seguintes perguntas sobre o tempo diário assistindo à TV para cada espectador.
- a) Qual é a probabilidade de que um espectador assista à TV durante 5 a 10 horas por dia?
- b) Por quantas horas um espectador deve assistir à TV para estar entre os 3% que mais assistem TV dentre todos os espectadores?
- c) Qual é a probabilidade de que um telespectador assista à TV por mais de 3 horas por dia?
- (Ex. 22 em Anderson et al, p. 259)

exercício 15

- O tempo necessário para concluir um exame final em determinado curso universitário está distribuído normalmente com uma média de 80 minutos e desvio padrão de 10 minutos. Responda às seguintes questões:
- a) Qual é a probabilidade de alguém concluir o exame em uma hora ou menos?
- b) Qual é a probabilidade de um estudante concluir o exame em mais de 60 minutos, porém, menos de 75 minutos?
- c) Suponha que a classe tenha 60 alunos e que a duração do exame seja de 90 minutos. Quantos estudantes você espera que não conseguirão concluir o exame no tempo determinado?
- (Ex. 23 em Anderson et al, p. 260)

exercício 16

- De acordo com a Sleep Foundation, o tempo médio de sono das pessoas à noite é de 6,8 horas (*Fortune*, 20 de março de 2006). Considere que o desvio padrão seja de 0,6 hora e que a distribuição de probabilidade seja normal.
- a) Qual é a probabilidade de que uma pessoa selecionada aleatoriamente durma mais de 8 horas?
- b) Qual é a probabilidade de que uma pessoa selecionada aleatoriamente durma 6 horas ou menos?
- c) Os médicos sugerem que se durma de 7 a 9 horas por noite. Qual porcentagem da população dorme durante o período recomendado?
- (Ex. 25 em Anderson et al, p. 260)

a distribuição normal

- Algumas propriedades da distribuição normal:
- Intervalo entre $-\sigma$ e σ contém $\sim$ 68,3% da probabilidade.
- Intervalo entre -2σ e 2σ contém $\sim$ 95,4% da probabilidade.
- Intervalo entre -3σ e 3σ contém $\sim$ 99,7% da probabilidade.
- (Desenhar no quadro.)
- Relevância disso? Essas propriedades nos permitem dispensar a “cola" em alguns casos.
- Exemplo: Qual a probabilidade de alguém ter mais de 1,88 de altura? Dados: $\mu = 170$cm e $\sigma = 6$cm
- $\sim$ 68,3% das pessoas terão entre 164cm e 176cm.
- $\sim$ 95,4% das pessoas terão entre 158cm e 182cm.
- $\sim$ 99,7% das pessoas terão entre 152cm e 188cm.
- R: Quem tem 1,88 ou mais está entre os 0,3% mais altos.