

Notas introdutórias em Probabilidade e Estatística

Luiz Renato Gonçalves Fontes
Instituto de Matemática e Estatística — USP

Prefácio

A motivação para estas notas é a apresentação de maneira unificadora do material de Estatística Descritiva e Probabilidade usualmente abordado em cursos introdutórios de Estatística em nível de graduação.

Os conceitos de distribuição de frequências (relativas) e de distribuição de probabilidades são centrais em, respectivamente, Estatística Descritiva e Probabilidade. Eles são análogos entre si, e admitem a mesma formulação matemática. Além disto, eles se relacionam direta e naturalmente no quadro da amostragem probabilística, conceito central em Inferência Estatística (que é o tema seguinte natural e usual em Estatística introdutória): quando fazemos amostragem casual simples numa população em que dada variável (populacional) está definida, tal variável medida na amostra se torna uma (ou mais, dependendo do tamanho da amostra) variável aleatória. Além disto, a distribuição de probabilidades desta variável aleatória pode ser identificada com a distribuição de frequências da variável populacional. Esta identidade é a base da Inferência Estatística.

Os dois primeiros capítulos tratam de pontos usuais de Estatística Descritiva, mas com ênfase nas distribuições de frequências, tratadas especificamente no primeiro capítulo, em que os casos discreto e contínuo são distinguidos; o primeiro em termos de funções de frequência, e o segundo em termos de funções densidade de frequência (aqui é inevitável recorrer a integração, que em muitos casos ainda não foi vista pela assistência; por isto lançamos mão, em apêndice, de uma definição informal e sucinta de integral). Distribuições multivariadas são abordadas apenas no caso discreto. No segundo capítulo, medidas de posição, dispersão e dependência são introduzidas, sempre com base nas distribuições de frequências (e pouca menção a dados brutos).

Os dois últimos capítulos versam sobre modelos probabilísticos e variáveis aleatórias, em abordagem por um lado mais tradicional, por outro destacando sempre o exemplo da amostragem probabilística em populações, fazendo conexão com o material dos capítulos iniciais, no espírito da discussão acima.

Sumário

1	Distribuições de freqüências	1
1.1	Representação gráfica de distribuições de freqüência	5
1.2	Função de densidade de freqüência	9
1.2.1	A densidade normal	13
1.3	Função de distribuição de freqüência	23
1.4	Distribuição de freqüências de mais de uma variável	28
1.4.1	Distribuição condicional e independência	31
2	Medidas descritivas para distribuições de freqüências	37
2.1	Medidas de posição	37
2.1.1	Média	38
2.1.2	Mediana	48
2.1.3	Quantis	53
2.1.4	Quartis e diagrama de caixa	55
2.2	Medidas de dispersão	57
2.2.1	Variância e Desvio-padrão	59
2.3	Uma medida de dependência entre variáveis	65
2.4	Associação linear entre duas variáveis	68
3	Probabilidade	80
3.1	Modelo probabilístico	81
3.2	Espaços equiprováveis	94
3.2.1	Passeio aleatório	99
3.3	Outros exemplos	103
3.4	Condicionamento e independência	106
3.4.1	Independência	115

4	Variáveis aleatórias	117
4.1	Distribuição de probabilidades de variáveis aleatórias	118
4.1.1	Variáveis aleatórias discretas	119
4.1.2	Variáveis aleatórias contínuas	120
4.1.3	Função de distribuição acumulada	122
4.1.4	Esperança	124
4.1.5	Variância	127
4.2	Modelos para variáveis aleatórias discretas	128
4.2.1	O modelo uniforme	128
4.2.2	O modelo de Bernoulli	129
4.2.3	O modelo binomial	129
4.2.4	O modelo hipergeométrico	134
4.2.5	O modelo geométrico	135
4.2.6	O modelo binomial negativo	136
4.2.7	O modelo de Poisson	137
4.3	Modelos para variáveis aleatórias contínuas	139
4.3.1	O modelo uniforme	139
4.3.2	O modelo exponencial	141
4.3.3	O modelo normal	143
4.4	Várias variáveis aleatórias	147
4.4.1	Condicionamento e independência	149
4.4.2	Covariância	152
4.4.3	Soma de variáveis aleatórias	154
A	Integral de uma função	160