EXERCÍCIOS DE ESTATÍSTICA

Jaime Fonseca

Daniel Torres

Vol. 2

2ª Edição
Revista e Corrigida

EDIÇÕES SÍLABO

EXERCÍCIOS
DE
ESTATÍSTICA

JAIME FONSECA
DANIEL TORRES

EDIÇÕES SÍLABO

É expressamente proibido reproduzir, no todo ou em parte, sob qualquer forma ou meio, **NOMEADAMENTE FOTOCÓPIA**, esta obra. As transgressões serão passíveis das penalizações previstas na legislação em vigor.

Visite a Sílabo na rede

www.silabo.pt

Editor: Manuel Robalo

FICHA TÉCNICA:

Título: Exercícios de Estatística – Vol. 2
Autores: Jaime Fonseca e Daniel Torres
© Edições Sílabo, Lda.
Capa: Pedro Mota

1ª Edição – Lisboa, Junho de 2002
2ª Edição – Lisboa, Outubro de 2014
Impressão e acabamentos: Europress, Lda.
Depósito Legal: 382188/14
ISBN: 978-972-618-774-5

EDIÇÕES SÍLABO, LDA.

R. Cidade de Manchester, 2
1170-100 Lisboa
Telfs.: 218130345
Fax: 218166719
e-mail: silabo@silabo.pt
www.silabo.pt

Índice

PREFÁCIO 7

Capítulo 6
ESTIMAÇÃO PONTUAL 9

Capítulo 7
ESTIMAÇÃO REGIONAL 65

Capítulo 8
TESTES DE HIPÓTESES 97

Capítulo 9
REGRESSÃO LINEAR 171

Capítulo 10
ANÁLISE DE VARIÂNCIA 223

BIBLIOGRAFIA 295

Prefácio

Na sequência do volume I surge agora o volume II de Exercícios de Estatística, necessariamente mais virado para a Inferência Estatística.

A Estatística pretende fundamentalmente concluir sobre a população ou populações em estudo, em aspectos variados como sejam, a título de exemplo:

— determinação de estimativas (pontuais ou regionais) de parâmetros desconhecidos;

— estabelecer relações de ordem entre parâmetros desconhecidos de duas ou mais populações;

— fazer Previsões para uma determinada variável.

Este volume II tenta acompanhar tais desígnios.

Usando uma perspectiva traçada no volume I, estamos seguros de que pode ser uma boa contribuição para o domínio desses aspectos.

Os autores

Capítulo 6

ESTIMAÇÃO PONTUAL

Exercício 1

Sabendo que certa população tem distribuição Binomial, com parâmetro p desconhecido, obtenha o estimador de máxima verosimilhança de p, com base numa amostra aleatória de dimensão n, $(X_1, X_2, ..., X_n)$, dela obtida. Proceda à análise do estimador.

Resolução:

Pode aplicar-se o método de máxima verosimilhança dado tratar-se de uma função regular segundo Cramer-Rao (é duplamente diferenciável e os limites de variação não dependem do parâmetro a estimar).

1. Obtenção do estimador

$$X \sim> f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} : \mu = np \;\; ; \;\; \sigma^2 = np(1-p)$$

$$L(p) = \prod_{i=1}^{n} \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i} =$$

$$= \prod_{i=1}^{n} \frac{n!}{x_i!\,(n-x_i)!}\, p^{\sum_{i=1}^{n} x_i} (1-p)^{\sum_{i=1}^{n} (n-x_i)}$$

Logaritmizando a função de verosimilhança, tem-se:

$$\log[L(p)] = \log\left[\prod_{i=1}^{n} \frac{n!}{x_i!\,(n-x_i)!}\right] + \log p^{\sum_{i=1}^{n} x_i} + \log(1-p)^{\sum_{i=1}^{n} (n-x_i)} =$$

$$= \sum_{i=1}^{n} \log\left[\frac{n!}{x_i!\,(n-x_i)!}\right] + \sum_{i=1}^{n} x_i \log p + \sum_{i=1}^{n} (n-x_i) \log(1-p) =$$

$$= \sum_{i=1}^{n} \log\left[\frac{n!}{x_i!\,(n-x_i)!}\right] + \sum_{i=1}^{n} x_i \log p + n^2 \log(1-p) - \sum_{i=1}^{n} x_i \log(1-p)$$

Derivando em ordem ao parâmetro,

$$\frac{d \log [L(p)]}{dp} = 0 + \sum_{i=1}^{n} x_i \frac{1}{p} + n^2 \frac{-1}{1-p} - \sum_{i=1}^{n} x_i \frac{-1}{1-p}$$

Igualando a zero, tem-se, após eliminar denominadores,

$$(1-p)\sum_{i=1}^{n} x_i - n^2 p + p\sum_{i=1}^{n} x_i = 0 \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \sum_{i=1}^{n} x_i - p\sum_{i=1}^{n} x_i - n^2 p + p\sum_{i=1}^{n} x_i = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{n} x_i - n^2 p = 0 \quad \Leftrightarrow$$

$$\Leftrightarrow \quad -n^2 p = -\sum_{i=1}^{n} x_i \quad \Leftrightarrow \quad p = \frac{\sum_{i=1}^{n} x_i}{n^2} \quad \Leftrightarrow \quad \hat{p} = \frac{\overline{x}}{n} \text{ (estimativa)}$$

2. Análise do estimador $\hat{p} = \dfrac{\overline{X}}{n}$

i) Não enviesamento

Um estimador diz-se não enviesado ou centrado, desde que $E[\hat{\theta}] = \theta$. Consequentemente,

$$E[\hat{p}] = E\left[\frac{\overline{X}}{n}\right] = E\left[\frac{\sum_{i=1}^{n} X_i}{n^2}\right] = \frac{1}{n^2} E\left[\sum_{i=1}^{n} X_i\right] =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} E[X_i] = \frac{1}{n^2}\sum_{i=1}^{n} np = \frac{1}{n^2} nnp = p$$

Assim, pode dizer-se que $\hat{\theta}$ é um estimador não enviesado para p.

ii) Coerência ou consistência

$$Var[\hat{p}] = Var\left[\frac{\overline{X}}{n}\right] = Var\left[\frac{\sum_{i=1}^{n} X_i}{n^2}\right] = \frac{1}{n^4} Var\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n^4} \sum_{i=1}^{n} Var[X_i]$$

$$= \frac{1}{n^4} \sum_{i=1}^{n} np(1-p) = \frac{1}{n^4} nnp(1-p) = \frac{p(1-p)}{n^2}.$$

Sendo $\lim_{n\to\infty} Var[\hat{p}] = \lim_{n\to\infty} \frac{p(1-p)}{n^2} = 0,$

Por ser não enviesado e por verificar a condição anterior, diz-se que $\frac{\overline{X}}{n}$ é um estimador coerente para p.

Exercício 2

a) Obtenha, através do método de máxima verosimilhança, com base na informação contida na amostra (4, 3, 0, 2, 1, 1), uma estimativa do parâmetro desconhecido da distribuição Poisson.

b) Dessa população extraiu-se uma amostra aleatória $(X_1, X_2, ..., X_5)$. Escolha, entre os dois estimadores seguintes, o mais adequado para estimar a média (μ) da população. Justifique adequadamente a sua opção.

$$T_1 = \overline{X} \; ; \qquad T_2 = \frac{1}{6}(2\ X_1 - X_2 + 3\ X_4 + 2\ X_5)$$

Resolução:

a) $X \sim> P(\lambda)$: $\qquad f_X(x) = \frac{\lambda^x}{x!}\ e^{-\lambda}; \qquad x = 0,1,2,\ldots$

Como a função é regular no sentido de Cramer-Rao (é duplamente diferenciável e os limites de variação não dependem do parâmetro a estimar) pode aplicar-se o método de máxima verosimilhança. Assim,

$$L(\lambda) = \prod_i \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \frac{\lambda^{\sum_i x_i}}{\prod_i x_i!} e^{-n\lambda}$$

$$\log L(\lambda) = \sum_i x_i \log \lambda - \sum_i \log x_i! - n\lambda$$

$$\frac{d \log L(\lambda)}{\partial \lambda} = \sum_i x_i \frac{1}{\lambda} - n$$

$$\frac{d \log L(\lambda)}{d\lambda} = 0 \quad \Leftrightarrow \quad \sum_i x_i \frac{1}{\lambda} - n = 0 \quad \Leftrightarrow \quad \sum_i x_i - n\lambda = 0 \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \hat{\lambda} = \frac{1}{n} \sum_i x_i = \bar{x} \qquad \text{(estimativa)}$$

$$\hat{\lambda} = \frac{1}{n} \sum_i X_i = \bar{X} \qquad \text{(estimador)}$$

Cálculo do valor da estimativa

$$\hat{\lambda} = \frac{1}{n} \sum_i x_i = \bar{x} = \frac{4 + 3 + 0 + 2 + 1 + 1}{6} = 1{,}83333$$

b) Estimadores Consistentes?

i) Análise do enviesamento

1. Caso de T_1

Para que seja um estimador não enviesado deverá ter-se $E[\hat{\lambda}] = \lambda$. Sendo

$$E[\hat{\lambda}] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} E\left[\sum_{i=1}^{n} X_i\right] =$$

$$= \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \frac{1}{n} \sum_{i=1}^{n} \lambda = \frac{1}{n} n\lambda = \lambda$$

T_1 constitui assim um estimador centrado ou não enviesado para $\mu = \lambda$

2. Caso de T_2

Identicamente, deverá ter-se $E[T_2] = \mu = \lambda$. Porque

$$E[T_2] = E\left[\frac{1}{6}(2X_1 - X_2 + 3X_4 + 2X_5)\right] =$$

$$= \frac{1}{6}(2E[X] - E[X_2] + 3E[X_4] + 2E[X_5]) =$$

$$= \frac{1}{6}(2\mu - \mu + 3\mu + 2\mu) = \frac{6\mu}{6} = \mu$$

T_2 constitui assim um estimador centrado ou não enviesado para $\mu = \lambda$

ii) Análise da consistência

1. Caso de T_1

Para que seja consistente ou coerente, deverá ter-se $\lim_{n \to +\infty} V[\hat{\lambda}] = 0$, desde que o estimador seja não enviesado. Ora

$$V[T_1] = V[\hat{\lambda}] = V\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} V\left[\sum_{i=1}^{n} X_i\right] =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} V[X_i] = \frac{1}{n^2}\sum_{i=1}^{n}\lambda = \frac{1}{n^2} n\lambda = \frac{\lambda}{n}$$

Sendo que $\lim \frac{\lambda}{n} \to 0$ e por se tratar de um estimador centrado podemos afirmar que o estimador é consistente.

2. Caso de T_2

$$\mathrm{Var}[T_2] = \mathrm{Var}\left[\frac{1}{6}(2X_1 - X_2 + 3X_4 + 2X_5)\right] =$$

$$= \frac{1}{36}(4Var[X_1] + Var[X_2] + 9Var[X_4] + 4Var[X_5]) =$$

$$= \frac{1}{36}(4\lambda + \lambda + 9\lambda + 4\lambda) = \frac{18\lambda}{36} = \frac{\lambda}{2}$$

Porque $\lim \frac{\lambda}{2} \neq 0$, pode concluir-se que o estimador T_2 não é consistente.

Em conclusão, face a estes resultados, poder-se-ía concluir nesta fase que o melhor estimador para a média da população é T_1.

Exercício 3

Obtenha através do método de máxima verosimilhança e da amostra que se segue, uma estimativa do parâmetro da distribuição exponencial.

4; 3; 2; 0; 3; 0

Resolução:

Obtenha-se em primeiro lugar a função de verosimilhança com base no produto das funções densidade exponencial.

Função densidade da variável com distribuição exponencial:

$$f_X(x) = \begin{cases} \lambda\ e^{-\lambda\ x} & ;\quad x \geq 0 \\ 0 & ;\quad x < 0 \end{cases}$$

$$L(\lambda) = \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} \lambda\ e^{-\lambda\ x_i} = \lambda^n\ e^{-\lambda \sum_{i=1}^{n} x_i}$$

Dado tratar-se de uma função regular no sentido de Cramer-Rao, pode aplicar-se o método de máxima verosimilhança.

Aplique-se a função logaritmo à função de verosimilhança para facilitar posteriormente a derivada.

$$\log\ L(\lambda) = \log \left(\lambda^n\ e^{-\lambda \sum_{i=1}^{n} x_i} \right) = \log\ \lambda^n + \log\ e^{-\lambda \sum_{i=1}^{n} x_i} =$$

$$= n\ \log\ \lambda - \lambda \sum_{i=1}^{n} x_i$$

Derive-se agora em ordem a λ e iguale-se a derivada a zero.

$$\frac{d \log L(\lambda)}{d \lambda} = n \frac{1}{\lambda} - \sum_{i=1}^{n} x_i = 0 \quad \Leftrightarrow$$

$$\Leftrightarrow \quad n \frac{1}{\lambda} = \sum_{i=1}^{n} x_i \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \frac{1}{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}$$

Logo com base na amostra dada tem-se o seguinte valor para a estimativa do parâmetro:

$$\hat{\lambda} = \frac{1}{\bar{x}} = \frac{6}{4 + 3 + 2 + 0 + 3 + 0} = \frac{6}{12} = 0{,}5$$

Exercício 4

Seja $(X_1, X_2, ..., X_n)$ uma amostra aleatória extraída de uma população com função densidade da forma:

$$f(x) = \begin{cases} \dfrac{x^2 \, e^{-x/\lambda}}{2\lambda^3} \; ; & x > 0 \quad , \quad \lambda > 0 \\ 0 \; ; & c.c. \end{cases}$$

a) Identifique o estimador de máxima verosimilhança do parâmetro desconhecido, com base na amostra aleatória.

b) Classifique o estimador, ao nível de enviesamento e consistência.

Resolução:

a) Obtenha-se em primeiro lugar a função de verosimilhança com base no produto das funções densidade.

$$L(\lambda) = \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} \frac{x_i^2 \, e^{-x_i / \lambda}}{2\lambda^3} = \left(\frac{1}{2\lambda^3} \right)^n \left(\prod_{i=1}^{n} x_i \right)^2 e^{-\sum_{i=1}^{n} x_i / \lambda}$$

Tratando-se de uma função regular segundo Cramer-Rao, aplique-se a função logaritmo à função de verosimilhança para facilitar posteriormente a derivada.

$$\log L(\lambda) = \log \left(\left(\frac{1}{2\lambda^3} \right)^n \left(\prod_{i=1}^{n} x_i \right)^2 e^{-\sum_{i=1}^{n} x_i / \lambda} \right) =$$

$$= -n \log(2\lambda^3) + 2 \sum_{i=1}^{n} \log x_i - \frac{\sum_{i=1}^{n} x_i}{\lambda}$$

Derive-se agora em ordem a λ e iguale-se a derivada a zero.

$$\frac{d \log L(\lambda)}{d \lambda} = -n \frac{3}{\lambda} + \frac{\sum_{i=1}^{n} x_i}{\lambda^2} = 0 \quad \Leftrightarrow$$

$$\Leftrightarrow \quad -3n\lambda + \sum_{i=1}^{n} x_i = 0 \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \hat{\lambda} = \frac{\overline{x}}{3} \quad \text{(estimativa)}$$

$$\Leftrightarrow \quad \hat{\lambda} = \frac{\overline{X}}{3} \quad \text{(estimador)}$$

b) Em função do resultado anterior ficamos a saber que o estimador correspondente é $\hat{\lambda} = \frac{\overline{X}}{3}$.

1. Enviesamento

$$E(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{n(3\lambda)}{n} = 3\lambda,$$

uma vez que $E[X] = 3\lambda$, tratando-se de uma distribuição Ga $(3, \lambda)$. Assim,

$$E(\hat{\lambda}) = \frac{1}{3} E(\overline{X}) = \lambda,$$

ou seja, $\hat{\lambda}$ constitui um estimador não enviesado ou centrado para λ.

2. Consistência

Atendendo a que $V[X] = 3\lambda^2$,

$$\lim_{n \to \infty} V(\hat{\lambda}) = \frac{1}{9} \lim_{n \to \infty} V(\overline{X}) = \frac{1}{9n^2} \sum_{i=1}^{n} V(X_i) = \lim_{n \to \infty} \frac{\lambda^2}{3n} = 0.$$

Conclui-se assim que $\hat{\lambda}$ é um estimador consistente para λ.

Exercício 5

Considere a variável aleatória X com a seguinte função densidade

$$f_X(x) = \frac{\delta + 3}{3} x^{\delta/3}; \qquad 0 \leq x \leq 1 \quad , \quad \alpha > 0$$

Estime com base numa amostra aleatória de tamanho n, através do método de máxima verosimilhança o parâmetro δ.

Resolução:

Como a função é regular no sentido de Cramer-Rao pode aplicar-se o método de máxima verosimilhança.

$$L(\delta) = \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \frac{\delta + 3}{3} x_i^{\delta/3} = \frac{(\delta + 3)^n}{3^n} \left(\prod_{i=1}^{n} x_i \right)^{\delta/3}$$

$$\log L(\delta) = \log \left(\frac{(\delta + 3)^n}{3^n} \left(\prod_{i=1}^{n} x_i \right)^{\delta/3} \right) =$$

$$= n \log(\delta + 3) - n \log 3 + \frac{\delta}{3} \sum_{i=1}^{n} \log x_i$$

$$\frac{d \log L(\delta)}{d\delta} = \frac{d}{d\delta} \left(n \log(\delta + 3) - n \log 3 + \frac{\delta}{3} \sum_{i=1}^{n} \log x_i \right) =$$

$$= \frac{n}{\delta + 3} + \frac{1}{3} \sum_{i=1}^{n} \log x_i$$

$$\frac{d \log L(\delta)}{d\delta} = 0 \quad \Leftrightarrow \quad \frac{n}{\delta + 3} + \frac{1}{3} \sum_{i=1}^{n} \log x_i = 0 \quad \Leftrightarrow$$

$$\Leftrightarrow \quad 3n + (\delta + 3) \sum_{i=1}^{n} \log x_i = 0 \quad \Leftrightarrow$$

$$\Leftrightarrow \quad (\delta + 3) \sum_{i=1}^{n} \log x_i = -3n \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \delta + 3 = \frac{-3n}{\sum_{i=1}^{n} \log x_i} \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \hat{\delta} = \frac{-3n}{\sum_{i=1}^{n} \log x_i} - 3$$

$\hat{\delta} = \dfrac{-3n}{\sum_{i=1}^{n} \log x_i} - 3$ (estimativa) e $\hat{\delta} = \dfrac{-3n}{\sum_{i=1}^{n} \log X_i} - 3$ (estimador)

DANIEL FERNANDES TORRES é licenciado em Economia (Universidade Lusíada, 1992) e mestre em Matemática Aplicada à Economia e à Gestão (Instituto Superior de Economia e Gestão Universidade Técnica de Lisboa, 1997). Possui uma pós-graduação em E-Business (Instituto Superior de Economia e Gestão, 2001). Obteve o DEA (*Diploma de Estudios Avanzados*) do programa de Doutoramento em *Sociedad de la Información y el Conocimiento* pela UPSAM (Universidade Pontifícia de Salamanca-Pólo de Madrid, 2005). É Técnico Oficial de Contas e membro efectivo da Ordem dos Economistas.

Desde 1992 tem exercido docência em várias universidades, entre as quais, Universidade Lusíada, Universidade Internacional, Instituto Superior de Gestão Bancária, Universidade Autónoma de Lisboa e Universidade Aberta.

Para além da docência, no início da sua carreira profissional estagiou e trabalhou como técnico no departamento financeiro de uma empresa de *factoring* de um grupo bancário, tendo vindo ao longo da última década a desenvolver, ao nível da gestão, actividade como administrador e gerente de empresas e ao nível contabilístico-financeiro, actividade como director administrativo-financeiro e como Técnico Oficial de Contas de várias empresas.

JAIME RAÚL SEIXAS FONSECA é licenciado em Engenharia Electrotécnica (Faculdade de Engenharia – Universidade do Porto, 1977) e mestre em Probabilidades e Estatística Computacional e Análise de Dados (Faculdade de Ciências – Universidade de Lisboa, 1988).

Leccionou em universidades públicas e privadas, entre as quais, Universidade Nova, Instituto Superior Técnico – Universidade Técnica de Lisboa, Universidade Lusíada, Universidade Moderna e Universidade Internacional. Paralelamente exerceu funções no Instituto Português da Qualidade.

Actualmente é docente no Instituto Superior de Ciências Sociais e Políticas – Universidade Técnica de Lisboa.

Baseado na longa experiência dos autores no ensino da estatística nos mais diversos cursos e escolas, esta obra apresenta um vasto conjunto de exercícios práticos resolvidos que permitirão aos alunos testar e validar os conhecimentos teóricos adquiridos.

Os estudantes de estatística encontrarão aqui as mais variadas aplicações, podendo seleccioná-las de entre os temas tratados, consoante a área de ensino e a profundidade da abordagem feita à matéria.

ISBN 978-972-618-774-5

9 789726 187745

146