Introdução à Probabilidade e Estatística

Universidade de Évora

Departamento de Matemática

Ano lectivo 2015/16

Ana Isabel Santos

Aulas 9 e 10

Intervalos de confiança

Estimação de parâmetros

Estimação pontual

Suponha-se que para uma certa população X, cuja função distribuição é $f(x;\theta)$, se desconhece o parâmetro caraterizador, θ, da distribuição. Portanto, a partir da informação que a amostra nos dá, pretende-se propor um valor para θ, que seja o melhor possível e que pertença ao espaço de resultados do parâmetro da população, Θ.

Donde, **um estimador de um parâmetro** da população é uma função da amostra aleatória $(X_1, X_2, \ldots, X_n)$, que não envolve parâmetros desconhecidos e o cujo contradomínio é Θ. Logo, um estimador **é uma variável aleatória**.

Estimação pontual

Assim, um **estimador pontual** é uma função que depende unicamente dos dados da amostra, que usualmente se representa por

$$\widehat{\theta} \equiv \widehat{\theta}(X_1, X_2, \ldots, X_n).$$

Donde, genericamente, podemos dizer que **qualquer estatística** é um estimador pontual.

Uma **estimativa** é um valor concreto assumido pelo estimador para uma amostra particular $(x_1, x_2, \ldots, x_n)$, ou seja, **é um número**.

Estimação pontual

Alguns estimadores pontuais mais usuais:

1. Estimador da média (μ)

$$\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n} \ ,$$

onde $(x_1, x_2, \ldots, x_n)$ é uma realização da amostra aleatória $(X_1, X_2, \ldots, X_n)$.

2. Estimador da variância (σ²)

$$S^2 = \sum_{i=1}^{n} \frac{\left(X_i - \overline{X}\right)^2}{n-1} \ .$$

Estimação intervalar

Na **estimação pontual** pretende-se obter um valor que, com base na informação amostral, seja o melhor valor para estimar um parâmetro populacional.

Na **estimação intervalar** constrói-se um intervalo de valores que contenha o verdadeiro valor do parâmetro, com um certo grau de certeza previamente estipulado.

Definição: Um **intervalo de confiança a (1 – α)x100%** para um parâmetro θ, onde **(1 – α)** é o **grau de confiança**, é um intervalo aleatório $\left]\, l_{\inf}, L_{\sup} \,\right[$ que se designa por **estimador intervalar**, tal que

$$P(l_{\inf} < \theta < L_{\sup}) = 1 - \alpha, \quad \alpha \in (0,1).$$

O **nível de significância** α deve ser um valor pequeno, para o grau de confiança seja elevado. Os valores usuais para a confiança são **90%, 95%** e **99%**.

Método de construção do I. C.

Construção do intervalo de confiança para o parâmetro θ:

A construção é feita através do método da variável pivot e consiste em:

- encontrar uma v.a. $Z(\vec{X},\theta)$ cuja distribuição seja independente de θ, onde $\vec{X} = (X_1, X_2, \ldots, X_n)$;

- fixar o grau de confiança $(1 - \alpha)$;

- obter dois pontos a e b (independentes de θ) tais que

$$P(a < Z(\vec{X},\theta) < b) = 1 - \alpha$$

e, posteriormente, na inversão destas desigualdades em ordem ao parâmetro θ.

Interpretação de Nível de Confiança

Considere-se, por exemplo, um nível de confiança de 95%.

1. Recolha-se uma amostra aleatória de dimensão n e construa-se o respetivo intervalo de confiança para o parâmetro θ.

2. Repitam-se estes passos para 100 amostras aleatórias da mesma população.

O que se pode dizer é que cerca de 95% desses intervalos contêm o valor do parâmetro θ, enquanto que os restantes 5% não contêm esse parâmetro.

Intervalo de confiança para a média

1. Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória independente e indentica-mente distribuída (a.a.i.i.d.) de uma **população Normal** $X \sim N(\mu;\sigma)$, com **variância** σ² **conhecida**. Dado $\alpha \in (0,1)$, pretende-se construir um intervalo de confiança (I. C.) a $(1-\alpha)\times 100\%$ para μ.

Como $\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, então $Z = \dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ é uma variável pivot para μ. Donde,

$$P(a < Z < b) = 1-\alpha \Rightarrow a = -z_{1-\frac{\alpha}{2}} \quad \text{e} \quad b = z_{1-\frac{\alpha}{2}}.$$

logo

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{1-\frac{\alpha}{2}}\right) = P\left(\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha,$$

Intervalo de confiança para a média

ou seja,

$$\left]\overline{X} - z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} ; \overline{X} + z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\right[$$

é o **intervalo de confiança a (1–α)×100% para μ**.

Observação: Quanto maior for a confiança com que se queira estimar o intervalo, maior será a amplitude deste.

À metade da amplitude de um intervalo de confiança designa-se por **margem de erro**.

Intervalo de confiança para a média

2. Seja $X_1, X_2, \ldots, X_n$ uma a. a. i. i. d. de uma **população Normal** $X \sim N(\mu;\sigma)$, com **variância** σ² **desconhecida**.

Dado $\alpha \in (0,1)$, pretende-se construir um I. C. a $(1-\alpha)\times 100\%$ para μ.

Neste caso, a variável pivot é

$$T = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}.$$

Donde,

$$P\left(\overline{X} - t_{n-1,1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{n-1,1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right) = 1-\alpha,$$

logo, o **intervalo de confiança a (1−α)×100% para μ** é dado por:

$$\left]\overline{X} - t_{n-1,1-\frac{\alpha}{2}} \times \frac{S}{\sqrt{n}};\ \overline{X} + t_{n-1,1-\frac{\alpha}{2}} \times \frac{S}{\sqrt{n}}\right[.$$

Intervalo de confiança para a média

3. Se a distribuição da população **não é Normal**, mas a dimensão da amostra é $n \geq 30$ e a **variância** σ^2 é **desconhecida**, então

$$\overline{X} \overset{\circ}{\sim} N\left(\mu, \frac{S}{\sqrt{n}}\right),$$

pelo que a variável pivot é

$$Z = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \overset{\circ}{\sim} N(0,1).$$

Logo, o **intervalo de confiança a (1–α)×100% para μ** é dado por:

$$\left]\overline{X} - z_{1-\frac{\alpha}{2}} \times \frac{S}{\sqrt{n}}; \overline{X} + z_{1-\frac{\alpha}{2}} \times \frac{S}{\sqrt{n}}\right[.$$

Margem de erro do I. C.

Margem de Erro do I. C. para a média - μ:

1. σ^2 conhecida: $z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$;

2. σ^2 desconhecida: $t_{n-1} \times \frac{S}{\sqrt{n}}$;

3. σ^2 desconhecida e $n \geq 30$: $z_{1-\frac{\alpha}{2}} \times \frac{S}{\sqrt{n}}$.

Intervalo de confiança para a variância

4. Seja $X_1, X_2, \ldots, X_n$ uma a. a. i. i. d. de uma **população Normal** $X \sim N(\mu;\sigma)$, com **variância** σ² desconhecida.

Dado $\alpha \in (0,1)$ pretende-se construir um I. C. a $(1-\alpha)\times 100\%$ para σ^2.

Neste caso, a variável pivot é

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Donde,

$$P\left(\frac{(n-1)S^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1;\frac{\alpha}{2}}}\right) = 1-\alpha,$$

logo, o **intervalo de confiança a $(1-\alpha)\times 100\%$ para σ^2** é dado por:

$$\left]\frac{(n-1)S^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}};\ \frac{(n-1)S^2}{\chi^2_{n-1;\frac{\alpha}{2}}}\right[.$$

Intervalo de confiança para a proporção

5. Seja $X_1, X_2, \ldots, X_n$ uma a. a. i. i. d. de uma **população de Bernoulli,** $Ber(p)$, com parâmetro p desconhecido.

Dado $\alpha \in (0,1)$ pretende-se construir um I. C. a $(1-\alpha)\times 100\%$ para p.

Neste caso, a variável pivot é

$$Z = \frac{\overline{p} - p}{\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}} \overset{\circ}{\sim} N(0,1), \quad \text{se } n > 30.$$

Donde,

$$P\left(\overline{p} - z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\overline{p}(1-\overline{p})}{n}} < p < \overline{p} + z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right) = 1-\alpha,$$

logo, o **intervalo de confiança a $(1-\alpha)\times 100\%$ para p²** é dado por:

$$\left]\overline{p} - z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\overline{p}(1-\overline{p})}{n}} ; \overline{p} + z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right[.$$

Intervalos de confiança

Exercício 1: A Inês afirma que, em média, os alunos da UÉ gastam 6.1 € por fim de semana em bebidas alcoólicas. Para testar a veracidade de tal afirmação, questionaram-se 100 alunos da UÉ tendo-se verificado que o gasto médio em bebidas alcoólicas ao fim de semana é de 3.4 €. Admita que o gasto em bebidas alcoólicas ao fim de semana segue uma distribuição normal com desvio padrão 1.8€.

a) Com 95% de confiança, concorda com a afirmação da Inês?

b) Sem efetuar cálculos, diga se mantém a sua opinião se considerar um grau de confiança de 90%. E se o grau de confiança for de 99%?

Intervalos de confiança

Exercício 2: O Rui é um jogador da equipa de futebol FCEstatística e pretende determinar um intervalo de confiança a 99% para o tempo médio em que participa nos jogos. Para tal, recolheu os dados relativos ao número de minutos que jogou em 10 jogos em que participou:

87, 76, 72, 86, 66, 77, 65, 81, 70 e 88.

Admita que o número de minutos que o Rui jogou segue uma distribuição Normal. Ajude o Rui a calcular o que pretende.

Intervalos de confiança

Exercício 2: Outputs do SPSS

Descriptive Statistics

	N	Mean	Std. Deviation
tempo de jogo	10	76,80	8,548

Intervalos de confiança

Exercício 3: Um conjunto de 40 condutores de camião, escolhidos aleatoriamente nas estradas nacionais, dispôs-se a participar numa experiência que tinha por objetivo medir os seus tempos de reação depois de almoço. A média e o desvio padrão dos tempos observados foram, respetivamente, 0.85 e 0.20 segundos. Admitindo que os tempos de reação seguem uma distribuição normal, determine:

a) o intervalo de confiança a 95% para o valor esperado do tempo de reação após o almoço.

b) o intervalo de confiança a 99% para a variância do tempo de reação após o almoço.

Intervalo de confiança

Exercício 4: Num trabalho realizado há já algum tempo concluiu-se que 62% dos passageiros que entram na estação A do metro tem como destino o centro da cidade. Esse valor tem vindo a ser utilizado em todos os estudos de transportes realizados desde então.

Tendo surgido dúvidas sobre a atualidade daquele valor, pois crê-se que tem vindo a diminuir, acompanhando o declínio do centro, realizou-se um inquérito naquela estação. Dos 240 passageiros inquiridos, 126 indicaram o centro como destino. Com base nestes resultados construa um intervalo de confiança a 90% para a percentagem de passageiros entrados em A e que saem no centro. O que pode concluir?

Intervalo de confiança para a diferença de médias

1. Considere-se duas amostras de dimensões n e m de duas populações independentes X e Y com valores médios μ_X e μ_Y e variâncias σ_X^2 e σ_Y^2. Admitamos que as duas populações são **Normais** com **variâncias conhecidas**. Então, o **intervalo de confiança a $(1-\alpha)\times 100\%$ para** $\mu_X - \mu_Y$ é dado por:

$$\left]\left(\overline{X}-\overline{Y}\right) - z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}\,;\,\left(\overline{X}-\overline{Y}\right) + z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}\right[.$$

Intervalo de confiança para a diferença de médias

2. Considere-se duas amostras de dimensões n e m de duas populações independentes X e Y com valores médios μ_X e μ_Y e variâncias σ_X^2 e σ_Y^2. Admitamos que as duas populações são **Normais** com **variâncias desconhecidas**, mas **iguais** $\sigma_X^2 = \sigma_Y^2$. Então, o **intervalo de confiança a (1–α)×100% para** $\mu_X - \mu_Y$ é dado por:

$$\left]\left(\overline{X} - \overline{Y}\right) - t_{n+m-2;1-\frac{\alpha}{2}} \times S^*; \left(\overline{X} - \overline{Y}\right) + t_{n+m-2;1-\frac{\alpha}{2}} \times S^*\right[,$$

onde

$$S^* = \sqrt{\frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}}\sqrt{\frac{1}{n} + \frac{1}{m}}.$$

Intervalo de confiança para a diferença de médias

3. Considere-se duas amostras de dimensões n e m de duas populações independentes X e Y com valores médios μ_X e μ_Y e variâncias σ_X^2 e σ_Y^2. Se as populações **não** são **Normais** e as **variâncias** são **desconhecidas**, mas **iguais**, $\sigma_X^2 = \sigma_Y^2$, e $n \geq 30$ e $m \geq 30$. Então, o **intervalo de confiança para** $\mu_X - \mu_Y$ **a (1–α)×100% de confiança** é dado por:

$$\left]\left(\overline{X} - \overline{Y}\right) - z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}} \,;\, \left(\overline{X} - \overline{Y}\right) + z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}\right[.$$

Intervalo de confiança para a diferença de médias

Exercício 5: Havendo indícios que o método de avaliação e as classificações finais atribuídas diferem fortemente entre em duas escolas (A e B), decidiu-se comprovar as estatísticas esta hipótese. Sabe-se os os desvios padrão são conhecidos, sendo 2,1 valores na escola A e 1,8 valores na escola B. Assim, recolheu-se uma amostra das notas dos testes dos alunos nas duas escolas e obtiveram-se os seguintes resultados:

Escola	n_i	$\overline{x}_i$
A	41	12,9
B	31	14,7

Recorrendo a um intervalo de confiança a 90%, diga se a classificação média da escola B é superior à da escola A. Justifique.