ALGUNS MÉTODOS DE ESTIMAÇÃO E
PREVISÃO EM POPULAÇÕES FINITAS E
DINÂMICAS

Lúcia Pereira Barroso

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA OBTENÇÃO DO GRAU DE MESTRE
EM
ESTATÍSTICA

ÁREA DE CONCENTRAÇÃO: ESTATÍSTICA
ORIENTADOR: PROF. DR. JOSEMAR RODRIGUES

- São Paulo, outubro de 1987 -

Aos meus pais

Adelaide e Manuel

AGRADECIMENTOS

Ao Prof. Dr. *Josemar Rodrigues*, pela sugestão do tema e pela dedicação com que me orientou.

Ao Prof. Dr. *Wilton de Oliveira Bussab*, pela colaboração e sugestões durante o trabalho.

Ao Prof. Dr. *Heleno Bolívarine*, pelos valiosos esclarecimentos.

Ao Prof. Dr. *Antonio Evaristo Teixeira Lanzana*, pelas discussões e incentivo.

À *Silvia Lopes de Paula Ferrari* e a *Rinaldo Bruno Schinazi*, pela amizade.

Aos demais professores do IME-USP, pela minha formação.

À *Antonia Soares*, pelo excelente trabalho de datilografia.

Finalmente, a todos aqueles que colaboraram para a realização deste trabalho.

Í N D I C E

INTRODUÇÃO	01
------------------	----

CAPÍTULO 1

1.1. Introdução	04
1.2. Objetivo do trabalho	07
1.3. Amostragem repetida	10

CAPÍTULO 2 - OBTENÇÃO DE ESTIMADORES ÓTIMOS ATRAVÉS DA TEORIA CLÁSSICA

2.1. Introdução	17
2.2. Procedimento para a obtenção de estimadores ótimos	19

CAPÍTULO 3 - OBTENÇÃO DE PREVISORES ÓTIMOS ATRAVÉS DA TEORIA DE SUPERPOPULAÇÕES

3.1. Introdução	24
3.2. Painel discreto	25
3.3. Aplicação I - caso discreto.....	32
3.4. Painel contínuo	38
3.5. Aplicação II - caso contínuo	41

CAPÍTULO 4 - OBTENÇÃO DE ESTIMADORES ÓTIMOS

ATRAVÉS DA TEORIA DE SUPERPOPULAÇÕES

4.1. Introdução	47
4.2. Procedimento para a obtenção de estimadores ótimos	49
4.3. Aplicação I - caso discreto (duas ocasiões)	52
4.4. Aplicação II - caso discreto (várias ocasiões)	66
4.5. Aplicação III - caso contínuo	69

CAPÍTULO 5 - OBTENÇÃO DE PREVISORES BAYESIANOS

ATRAVÉS DO MODELO DE FILTRO DE KALMAN

5.1. Introdução	73
5.2. Modelo de Filtro de Kalman com variâncias conhecidas	74
5.3. Modelo de Filtro de Kalman com variâncias desconhecidas	84

CAPÍTULO 6 - OBTENÇÃO DE ESTIMADORES ÓTIMOS

ATRAVÉS DA TEORIA DE MÍNIMOS QUADRADOS

6.1. Introdução	90
6.2. Obtenção dos estimadores, considerando μ conhecida	92
6.3. Obtenção dos estimadores, considerando μ desconhecida	101

CAPÍTULO 7 - ILUSTRAÇÕES NUMÉRICAS

7.1. Introdução	104
7.2. Um exemplo com dados reais	104
7.3. Um exemplo com dados simulados	111

APÊNDICE	118
----------------	-----

REFERÊNCIAS BIBLIOGRÁFICAS	136
----------------------------------	-----

oOo

INTRODUÇÃO

A maioria das análises estatísticas em Populações Finitas se baseia em uma amostra selecionada em determinado instante. Se a população não é dinâmica, isto é, as características populacionais não variam com o tempo, as informações obtidas nesse instante são válidas ao longo do tempo. Por outro lado, se a população é dinâmica, essas informações valem somente para o instante em que a amostra foi selecionada.

Entretanto, muitos problemas práticos de interesse, principalmente na área social e econômica, como por exemplo, opinião pública e desemprego, referem-se a populações dinâmicas. Na área de inferência estatística em Populações Finitas, esses problemas não tem sido claramente formulados, o que se pode verificar pelo pequeno número de artigos publicados nesse sentido.

Assim sendo, a finalidade deste trabalho é colocar de forma genérica e em uma única linguagem, os procedimentos da Teoria Clássica e da Teoria de Superpopulações para populações dinâmicas.

Objetivamente, o trabalho procurará responder às seguintes perguntas:

- 1) Como obter os dados em uma população dinâmica? (amostra - gem repetida)
- 2) Como extrair informações desses dados? (estimação ou previ são)

A apresentação será feita em 7 capítulos. No Capí - tulo 1 será definida a função paramétrica populacional, θ , de in - teresse. Além disso serão comentados alguns tipos de amostras ao longo do tempo.

No Capítulo 2 será apresentado um método de obtenção de estimadores ótimos para θ , através da Teoria Clássica, se - gundo um critério de otimalidade fixado.

No Capítulo 3, a abordagem será a de Superpopulações e o previsor ótimo para θ será obtido através do uso do Teore - ma de Gauss-Markov para Populações Finitas.

No Capítulo 4, ainda sob a abordagem da Teoria de Su - perpopulações, serão obtidos estimadores ótimos para $E(\theta)$, se - gundo um critério de otimalidade.

Já no Capítulo 5, o enfoque será Bayesiano e o previ - sor de θ será obtido através do Modelo de Filtro de Kalman.

No Capítulo 6 será apresentado, para a obtenção do es - timador ótimo para θ , o Método dos Mínimos Quadrados, quando os parâmetros são variáveis aleatórias.

Finalmente, no Capítulo 7, serão apresentadas algu - mas ilustrações numéricas de alguns estimadores (previsores).

Para isso serão usados um conjunto de dados reais e um conjunto de dados simulados.

oOo

CAPÍTULO 1

1.1. Introdução

A fim de acompanhar determinada característica de uma população, durante várias ocasiões ou durante certo período de tempo, serão definidas duas populações: $P_1(t) = \{1, 2, \dots, N(t)\}$, que representa a população de elementos no instante t , e $P_2(t)$, que pode ser discreta ($P_2(T) = \{0, 1, 2, \dots, T\}$) ou contínua, ($P_2(T) = \{t: 0 \leq t \leq T\}$), que representa a população dos instantes de tempo.

As seguintes suposições serão feitas:

- a) Não há entradas ou saídas de elementos na população $P_1(t)$. Consequentemente, pode-se representar $P_1(t)$ por $P_1(t) = \{1, 2, \dots, N\}$, onde N é o tamanho da população, para qualquer $t \in P_2(T)$.
- b) N e T são conhecidos.

A variável de interesse para o k -ésimo elemento de $P_1(t)$ será representada por $Y_k(t)$, $1 \leq k \leq N$, e o vetor populacional no instante t , por

$$\underline{Y}'(t) = (Y_1(t), Y_2(t), \dots, Y_N(t)) , t \in P_2(T).$$

Aqui, algumas considerações a respeito de $\underline{Y}'(t)$ são importantes:

- 1) Se $\underline{Y}'(t)$ está associado a um vetor de números reais fixos, desconhecidos, então o problema está sendo analisado sob o ponto de vista da Teoria Clássica. Neste caso, diz-se que a população é fixa e a fonte de aleatorização é o método pelo qual as unidades da amostra são selecionadas.
- 2) Se $\underline{Y}'(t)$ é um vetor de variáveis aleatórias e os valores observados são vistos como realizações dessas variáveis, o problema passa a ser encarado sob a Teoria de Superpopulações. (Cassel, Särndal, Wretman - 1977).

Para ilustrar essas duas abordagens, Cassel/Särndal/Wretman (1977) citam um exemplo, onde as unidades populacionais são fazendas e a característica de interesse é a produção de trigo em determinado ano. Sob o enfoque da Teoria Clássica, supõe-se que a cada fazenda na população, corresponde um número real fixo, porém desconhecido, que representa a produção de trigo naquele particular ano.

No outro enfoque, da Teoria de Superpopulações, as produções das fazendas são tratadas como números gerados sob um modelo probabilístico. Em geral, esses modelos incorporam informações auxiliares. No exemplo, os autores sugerem uma re

lação linear, sendo a produção de trigo, fora um erro de valor esperado zero, proporcional ao tamanho da fazenda em acres, x_k . Além disso, eles consideram que o fator de proporcionalidade β (desconhecido) é comum a todas as fazendas.

- 3) Se $T = 0$, $P_1(t)$ está fixada num único instante. Isto significa que há interesse na característica, somente naquele instante fixado e então, uma única amostra nesse instante seria suficiente para se obter informação a respeito da população.
- 4) Se $T > 0$ e $Y_k(t) = c_k$ (constante através do tempo) então a população $P_1(t)$ não é dinâmica. Neste caso, a característica associada a qualquer elemento da população $P_1(t)$ não varia com o tempo e, também aqui, uma única amostra em um só instante de tempo é suficiente para o estudo, até o final do período de interesse.
- 5) Se $T > 0$ e $Y_k(t)$ varia com t , então $P_1(t)$ é dinâmica. Se há interesse nessa característica dinâmica, os dados devem ser coletados em vários pontos através do tempo, ou durante todo o intervalo de tempo.

É importante ressaltar que, quando a população é dinâmica, as variações podem ser devidas a dois fatores: mudança da característica do indivíduo ou mudança na composição da população. Como está sendo feita a suposição de que os elementos de $P_1(t)$ são os mesmos para qualquer $t \in P_2(T)$ (suposição

a), o segundo fator de variação foi eliminado.

1.2. Objetivo do Trabalho

As seguintes definições serão úteis para a formulação do objetivo do trabalho:

Definição 1.1: Será chamada de função paramétrica populacional no instante t , representada por $\theta(t)$, qualquer função dos elementos do vetor $\underline{Y}(t)$, ou seja,

$$\theta(t) = \theta(\underline{Y}'(t)) \quad , \quad \forall t \in P_2(T)$$

Exemplos: (i) Média Populacional: $\theta(t) = \frac{1}{N} \sum_{k=1}^N Y_k(t) = \bar{Y}(t)$

(ii) Total Populacional: $\theta(t) = \sum_{k=1}^N Y_k(t)$

(iii) Variância Populacional: $\theta(t) = \frac{1}{N} \sum_{k=1}^N (Y_k(t) - \bar{Y}(t))^2$.

Neste trabalho serão examinadas somente funções paramétricas lineares de $Y_k(t)$.

Definição 1.2: A característica populacional de interesse, associada à população $P_1(t)$, para $t \in P_2(T)$, será definida por:

$$\theta = \int_0^T c(t) \theta(t) . dF(t) \quad , \quad (1.1)$$

onde $c(t)$ é uma função conhecida, exemplificada a seguir, e $F(t)$ é uma distribuição de probabilidade definida em $P_2(T)$.

A característica populacional θ expressa o comportamento da população ao longo do tempo.

O objetivo do trabalho consiste em construir um estimador (previsor) ótimo $\hat{\theta}$ para θ , dadas as funções $c(t)$, $\theta(t)$ e $F(t)$ e um esquema amostral. O estimador (previsor) $\hat{\theta}$ deverá ser ótimo para θ , segundo algum critério a ser determinado pelo pesquisador. Além disso, sob o enfoque da Teoria de Superpopulações, será obtido um estimador ótimo para $E[\theta]$.

As funções $c(t)$, $\theta(t)$ e $F(t)$ serão especificadas de acordo com o objetivo do estudo.

Alguns exemplos serão dados abaixo, para a média no instante t , caso em que $\theta(t) = \frac{1}{N} \mathbf{1}'_N \mathbf{y}(t)$, onde $\mathbf{1}'_N$ é o vetor linha de tamanho N , com todos os elementos iguais a 1.

Exemplo 1.1: Quando a característica da população, $y_k(t)$, muda rapidamente com o passar do tempo, pode haver interesse em se estimar o valor médio, na ocasião mais recente (T). Neste caso,

$$c(t) = 1, \quad \forall t \in P_2(T),$$
$$F(t) = \begin{cases} 0, & p|t < T \\ 1, & p|t = T. \end{cases}$$

Pelas propriedades da Integral (resultado 1 do apên-

dice),

$$\theta = \theta(T) = \frac{1}{N} \sum_{N} Y(t) .$$

Exemplo 1.2: Quando as mudanças na característica da população são lentas, a avaliação da média geral da população é útil e, neste caso: $c(t) = 1$, $F(t)$ é a distribuição uniforme em $[0, T]$, $\forall t \in P_2(T)$ e então

$$\theta = \int_0^T 1 \cdot \theta(t) \frac{1}{T} dt = \frac{1}{T} \int_0^T \theta(t) dt .$$

Quando $N = 1$, este parâmetro foi estudado com detalhes por Bartlett (1986).

Exemplo 1.3: Pode haver interesse em medir a variação entre duas ocasiões, por exemplo, T e $T-1$, ou seja, $\theta(T) - \theta(T-1)$. Nesse caso, $F(t)$ será dada por:

t	0	1		T-1	T
dF(t)	P_0	P_1		P_{T-1}	P_T

e $c(t)$ por

$$\begin{cases} c(T) = 1/P_T \\ c(T-1) = -1/P_{T-1} \\ c(t) = 0 \text{ para } t \neq T-1 \text{ e } t \neq T \end{cases}$$

Novamente, pelo resultado 1 do apêndice,

$$\theta = \sum_{t=0}^T c(t) \theta(t) p_t = \theta(t) - \theta(T-1).$$

A fim de estimar (prever) θ , duas alternativas de planejamento serão citadas: amostrar os elementos de $P_1(t)$ para todos os instantes de $P_2(T)$, ou amostrar os elementos de $P_1(t)$ para alguns instantes de $P_2(T)$.

Se $P_2(T)$ é uma população discreta, então em ambas alternativas, os dados serão discretos no tempo. Por outro lado, se $P_2(T)$ é contínua, então os dados podem ser discretos ou contínuos no tempo.

Como $c(t)$ e $F(t)$ são funções especificadas, a única parte desconhecida na expressão $\theta = \int_0^T c(t) \theta(t) dF(t)$ será $\theta(t)$ e então, o estimador natural de θ será $\hat{\theta} = \int_0^T c(t) \hat{\theta}(t) dF(t)$, onde $\hat{\theta}(t)$ é um estimador (previsor) de $\theta(t)$, escolhido segundo algum critério de otimalidade a ser adotado.

1.3. Amostragem repetida

Nesta seção serão introduzidos alguns esquemas amostrais que permitirão obter os dados em uma população dinâmica.

Há várias maneiras de se coletar dados através do tempo. No caso em que os dados são discretos, três planos para seleção serão aqui citados: amostras independentes, painel (completo e incompleto) e painel rotativo. No caso contínuo,

será citado o painel, completo e incompleto.

Dados discretos

Amostras Independentes - As amostras são selecionadas da população em várias ocasiões no tempo, independentemente umas das outras.

O conjunto dos dados das amostras independentes, para os vários instantes, será denotado por $\mathcal{D}_{IN}$, ou seja,

$$\mathcal{D}_{IN} = \{ (k, Y_k(t)) , k \in S(t) , t \in \{0, 1, 2, \dots, T\} \},$$

onde $S(t)$ é uma amostra de tamanho n obtida no instante t , segundo um planejamento amostral adotado pelo pesquisador.

É válido salientar que, eventualmente, um mesmo elemento da população pode ser amostrado em mais de uma ocasião, mas isso não é uma exigência do plano amostral.

Painel - Uma amostra é selecionada da população no primeiro instante, e depois os dados são coletados, dessa mesma amostra, para várias ocasiões. Esses dados serão denominados Painel Completo, e o conjunto de dados denotado por $\mathcal{D}_{PC}$, ou seja,

$$\mathcal{D}_{PC} = \{ (k, Y_k(t)) , k \in S(t) = S , t \in \{0, 1, 2, \dots, T\} \}.$$

Às vezes, apesar do indivíduo ser amostrado várias vezes, ele fornece a resposta somente para algumas ocasiões e, então algumas das observações são perdidas. Nesse caso, o con

junto de dados será denominado painel incompleto, denotado por $\mathcal{D}_{PI}$, isto é,

$$\mathcal{D}_{PI} = \{ (k, Y_k(t)) , k \in S(t) , t \in \{0, 1, \dots, T\} \} ,$$

onde $S(0) = S$ e $S(j) \subset S$, $j = 1, 2, \dots, T$ e existe i tal que $S(i) \neq S$.

Painel Rotativo - Uma amostra é selecionada em um determinado instante e, na ocasião seguinte, uma parte dessa amostra é mantida e o restante é substituído por novos elementos, através de novo sorteio. Este conjunto de dados será denotado por $\mathcal{D}_{PR}$,

$$\mathcal{D}_{PR} = \{ (k, Y_k(t)) , k \in S(t) , t \in \{0, 1, \dots, T\} \} ,$$

onde

$$S(j) = \{k_1, k_2, \dots, k_n\} ,$$

$$S(j-1) = \{k_1, k_2, \dots, k_m, k_{n+1}, k_{n+2}, \dots, k_{n+u}\} ,$$

$$j = 1, 2, \dots, T \quad \text{e} \quad m + u = n$$

Neste caso, m elementos foram amostrados nas duas ocasiões, j e $j-1$ e $u = n-m$ elementos, somente em uma delas.

Para ilustrar esse tipo de amostra, suponha que a população $P_1(t)$ tenha 6 unidades e que uma amostra de 3 unidades é retirada da população, para cada instante, sendo que, entre as amostras dos instantes j e $j-1$ existem 2 unidades comuns.

Uma esquematização disso pode ser a seguinte:

instante \ unidade	1	2	3	4	5	6	7
1	X				X	X	X
2	X	X				X	X
3	X	X	X				X
4		X	X	X			
5			X	X	X		
6				X	X	X	

No instante 1, a amostra é composta pelas unidades 1, 2 e 3. No instante 2, a amostra consta das unidades 2, 3 e 4, ou seja, existem duas unidades (2 e 3) comuns às amostras nesses dois instantes.

Dados contínuos

No caso contínuo será usado somente o painel, ou seja, uma amostra é selecionada no primeiro instante, e sua trajetória é observada durante todo o intervalo de tempo.

Isto poderia ser representado pelo gráfico da figura 1.1, onde cada linha é a trajetória do indivíduo durante o intervalo de tempo $[0, T]$. O conjunto das trajetórias será denotado por $\mathcal{D}_{PC}$,

Figura 1.1 - exemplo de painel contínuo completo

$$\mathcal{D}_{PC} = \{ (k, Y_k(t)) , k \in S(t) = S , t \in P_2(T) = [0, T] \} .$$

Também no caso contínuo, poderia ocorrer que uma parte dos indivíduos amostrados deixassem, a partir de determinado instante, de responder à pesquisa.

Uma representação disso poderia ser dada pelo gráfico da figura 1.2, onde a trajetória daqueles que deixaram de responder seria incompleta. Este conjunto de dados será denominado painel contínuo incompleto, denotado por $\mathcal{D}_{PI}$,

$$\mathcal{D}_{PI} = \{ (k, Y_k(t)) , k \in S(t) , t \in P_2(T) = [0, T] \}$$

onde $S(0) = S$ e $S(j) \subset S$, $j \in (0, T]$, e existe i tal que para $j \in (i, T]$, $S(j) \neq S$.

Cada um desses tipos de planos apresentam vantagens e desvantagens e algumas delas, tiradas de Cochran (1963) e de

Duncan/Kalton (1987), serão apresentadas a seguir.

Figura 1.2 - exemplo de painel contínuo incompleto

Como vantagem do painel, pode-se citar o seu maior potencial analítico em relação à amostra independente, porque é possível coletar um número maior de informações, divididas pelas várias ocasiões. Pode-se também detectar erros nas respostas com maior facilidade, uma vez que um informante pode fornecer a mesma informação mais de uma vez.

Além disso, a experiência tem mostrado que os informantes têm mais boa vontade e fornecem dados mais aprimorados na segunda entrevista, principalmente quando as informações são técnicas ou confidenciais.

Outras vantagens do painel são os custos mais baixos nas obtenções futuras de informações e a possibilidade de acumular informações para estudar populações raras.

Por outro lado, os principais problemas apresentados pelo painel é que os informantes podem ficar cansados de fornece

cer sempre o mesmo tipo de informação e, desse modo, aumentar o número de não respostas.

Os informantes podem, ainda, com o tempo, tornarem-se influenciados pelo pesquisador e assim, serem menos representativos da população.

Ainda como desvantagem, o painel não leva em conta que a composição da população pode estar mudando e desconsidera a introdução de novos elementos na população.

Pode ainda ocorrer a dificuldade de se entrar em contato com os informantes periodicamente, uma vez que pode haver mudança de endereço ou mortalidade do informante. Além disso, pode ocorrer dificuldade em seguir com o painel quando há divisão da unidade observacional, por exemplo, família.

Em todos esses casos, o que é vantagem para o painel é desvantagem para a amostra independente e vice-versa.

Pode-se ainda acrescentar que a amostra independente torna os cálculos mais simples do que no caso do painel.

Algumas vezes o painel rotativo apresenta melhores resultados do que o painel ou a amostra independente. Na verdade, a escolha do plano depende em parte do objetivo do problema em estudo.

Duncan/Kalton (1987) listaram alguns objetivos e apresentaram, para cada um, qual o plano amostral mais adequado.

CAPÍTULO 2

OBTENÇÃO DE ESTIMADORES ÓTIMOS ATRAVÉS DA TEORIA CLÁSSICA

2.1. Introdução

A análise clássica dos dados obtidos em amostragem repetida ao longo do tempo, em uma população finita, foi introduzida por Patterson em 1950. Dentro de um contexto mais geral do que Patterson, o objetivo, neste capítulo, é estimar θ através do uso da abordagem clássica. Como $c(t)$ e $F(t)$ são funções especificadas, o problema se resume em estimar $\theta(t)$ e neste caso, $\hat{\theta}$ será dado por

$$\hat{\theta} = \int_0^T c(t) \hat{\theta}(t) dF(t) ,$$

onde

$$\hat{\theta}(t) \in u_1(t) ,$$

$$u_1(t) = \{\hat{\theta}(t) : \hat{\theta}(t) = q\hat{\theta}(\mathcal{D}) + (1-q)\hat{\theta}(d(t)) \text{ e } 0 \leq q \leq 1\},$$

$\mathcal{D}$ é conjunto dos dados obtidos segundo um dos planos citados no capítulo anterior e $d(t) = \{(k, y_k(t)), k \in S(t)\}$. $u_1(t)$ é a classe dos estimadores de $\theta(t)$, compostos pela média ponderada de dois outros estimadores de $\theta(t)$: $\hat{\theta}(\mathcal{D})$, que leva em consideração todo o conjunto de dados até o instante t e $\hat{\theta}(d(t))$, que considera somente os dados do instante t .

A escolha de um estimador pertencente à classe $u_1(t)$ foi motivada por um estimador desse tipo, apresentado por Cochran (1963). O problema de se obter $\hat{\theta}(t) \in u_1(t)$ é conhecido na literatura de séries temporais como problema de filtração (ver Kalman, 1960).

Usando-se as propriedades da Integral (resultado 2 do apêndice), $\hat{\theta}$ pode ser escrito da seguinte forma:

$$\hat{\theta} = q \int_0^T c(t) \hat{\theta}(d) dF(t) + (1-q) \int_0^T c(t) \hat{\theta}(d(t)) dF(t) \quad (2.1)$$

Critério para otimalidade

O critério para obtenção do estimador será o seguinte: fixados os estimadores $\hat{\theta}(d)$ e $\hat{\theta}(d(t))$, será encontrado o valor de "q", que torna $\hat{\theta}$ ótimo para θ . A definição de estimador ótimo será dada adiante (definição 2.1).

Notação

Dados os estimadores $\hat{\theta}(d)$ e $\hat{\theta}(d(t))$, u_2 é a classe dos estimadores de θ , não viciados, do tipo (2.1), em que $\hat{\theta}(t)$ pertence à classe $u_1(t)$, isto é,

$$u_2 = \{\hat{\theta} : \hat{\theta}(t) \in u_1(t) \text{ e } E_p(\hat{\theta}) = \theta\},$$

onde E_p indica a esperança de $\hat{\theta}$, com relação ao planejamento (Rao - 1973).

Definição 2.1 - O estimador $\hat{\theta}^*$ é ótimo para θ , segundo esse critério, se e somente se é não viciado para θ e sua variância é menor ou igual à variância de qualquer outro estimador não viciado para θ , ou seja, satisfaz as duas condições abaixo:

- i) $\hat{\theta}^* \in \mathcal{U}_2$
- ii) $\text{Var}[\hat{\theta}^*] \leq \text{Var}[\hat{\theta}], \forall \hat{\theta} \in \mathcal{U}_2$

2.2. Procedimento para a obtenção de estimadores ótimos

Teorema 2.1 - Se $\hat{\theta}(\mathcal{D})$ e $\hat{\theta}(d(t))$ são não viciados para $\theta(t)$, então $\hat{\theta}$ é não viciado para θ .

Prova:

$$\begin{aligned}
 E_p(\hat{\theta}) &= E_p \left[q \int_0^T c(t) \hat{\theta}(\mathcal{D}) dF(t) + (1-q) \int_0^T c(t) \hat{\theta}(d(t)) dF(t) \right] = \\
 &= q \int_0^T c(t) E_p[\hat{\theta}(\mathcal{D})] dF(t) + (1-q) \int_0^T c(t) E_p[\hat{\theta}(d(t))] dF(t) = \\
 &= q \int_0^T c(t) \theta(t) dF(t) + (1-q) \int_0^T c(t) \theta(t) dF(t) = \\
 &= \int_0^T c(t) \theta(t) dF(t) = \theta .
 \end{aligned}$$

Teorema 2.2 - Se $\hat{\theta}(\mathcal{D})$ e $\hat{\theta}(d(t))$ são não viciados para $\theta(t)$ e independentes para qualquer $t \in P_2(T)$, então o estimador ótimo para θ , segundo o critério acima, é dado por

$$\hat{\theta}^* = q^* \int_0^T c(t) \hat{\theta}(\varnothing) dF(t) + (1-q^*) \int_0^T c(t) \hat{\theta}(d(t)) dF(t), \quad (2.2)$$

onde

$$q^* = \frac{\text{Var} \left(\int_0^T c(t) \hat{\theta}(d(t)) dF(t) \right)}{\text{Var} \left(\int_0^T c(t) \hat{\theta}(d(t)) dF(t) \right) + \text{Var} \left(\int_0^T c(t) \hat{\theta}(\varnothing) dF(t) \right)} \quad (2.3)$$

Prova: Seja

$$a = \int_0^T c(t) \hat{\theta}(\varnothing) dF(t) \quad \text{e} \quad b = \int_0^T c(t) \hat{\theta}(d(t)) dF(t)$$

i) Como $\hat{\theta}(\varnothing)$ e $\hat{\theta}(d(t))$ são, por hipótese, não viciados para $\theta(t)$, pelo teorema 2.1, $\hat{\theta}^*$ é não viciado para θ , logo $\hat{\theta}^* \in \mathcal{U}_2$.

ii) Seja $\hat{\theta} \in \mathcal{U}_2$

$$\begin{aligned} \text{Var}[\hat{\theta}] &= \text{Var} \left[q \int_0^T c(t) \hat{\theta}(\varnothing) dF(t) + (1-q) \int_0^T c(t) \hat{\theta}(d(t)) dF(t) \right] = \\ &= q^2 \text{Var} \left(\int_0^T c(t) \hat{\theta}(\varnothing) dF(t) \right) + (1-q)^2 \text{Var} \left(\int_0^T c(t) \hat{\theta}(d(t)) dF(t) \right) = \\ &= q^2 \text{Var}(a) + (1-q)^2 \text{Var}(b). \end{aligned}$$

Minimizando $\text{Var}[\hat{\theta}]$ com relação a "q",

$$\frac{d \text{Var}[\hat{\theta}]}{dq} = 2q \text{Var}(a) - 2(1-q) \text{Var}(b) = 0$$

$$q \text{ Var}(a) + q \text{ Var}(b) = \text{Var}(b) ; q = \frac{\text{Var}(b)}{\text{Var}(a) + \text{Var}(b)} ,$$

ou seja, o valor de "q" que minimiza $\text{Var}[\hat{\theta}]$ é q^* .

Pode-se citar como um caso particular de (2.2), o estimador obtido por Cochran (1963), onde uma amostra aleatória simples de n elementos foi sorteada da população, em duas ocasiões. O parâmetro θ corresponde à média da população na segunda ocasião.

Nesse exemplo foi usado um painel rotativo, onde m elementos foram mantidos para as duas ocasiões. O melhor estimamador encontrado é a combinação de dois estimadores independentes: um, que leva em conta a amostra na primeira ocasião mais a parte da amostra mantida para a segunda ocasião e outro, que considera somente a parte nova da amostra, na segunda ocasião.

O fator de ponderação é a proporção correspondente ao inverso da variância de cada um dos estimadores, como também ocorre no teorema 2.2, uma vez que q^* pode ser escrito como

$$q^* = \frac{\frac{1}{\text{Var}(a)}}{\frac{1}{\text{Var}(a)} + \frac{1}{\text{Var}(b)}} .$$

Além disso, Cochran (1963) generalizou esse estima-mador composto, para mais do que duas ocasiões.

Outro estimador similar a $\hat{\theta}^*$ foi dado por Graham/Rao (1964). Eles analisaram dois casos: a média da população em determinado instante, designado instante "0", e a diferença en

tre as médias para dois instantes consecutivos, "0" e "-1".

Para o primeiro caso, o estimador proposto, obtido recursivamente, foi dado por:

$$\bar{x}'_0 = Q(\bar{x}'_{-1} + \bar{x}_{0,-1} - \bar{x}_{-1,0}) + (1-Q)\bar{x}_0, \quad 0 \leq Q < 1,$$

onde

$\bar{x}'_{-1}$ indica o mesmo estimador na etapa anterior,

$\bar{x}_{0,-1}$ é a média, no instante "0", dos elementos mantidos da etapa "-1",

$\bar{x}_{-1,0}$ é a média, no instante "-1", dos elementos mantidos para a etapa "0", e

$\bar{x}_0$ é a média de toda a amostra, no instante "0".

Também neste caso, pode-se observar que o estimador composto é a média ponderada entre dois estimadores: um, que considera as amostras nos dois instantes, "0" e "-1", e outro, que considera somente a amostra no instante "0".

Uma diferença, que vale a pena destacar, entre este estimador e aquele proposto por Cochran (1963) é que, neste caso, os componentes não são independentes.

Para a diferença, o estimador proposto foi

$$d'_0 = \bar{x}'_0 - \bar{x}'_{-1}.$$

Nesse artigo, Graham/Rao (1964) ainda apresentaram explicitamente, as variâncias de $\bar{x}'_0$ e d'_0 e, para algumas situações especificadas, calcularam a porcentagem de ganho em e-

ficiência, com relação ao estimador usual, $\bar{x}_0$ e $\bar{x}_0 - \bar{x}_{-1}$, respectivamente.

oOo

CAPÍTULO 3

OBTENÇÃO DE PREVISORES ÓTIMOS ATRAVÉS DA TEORIA DE SUPERPOPULAÇÕES

3.1. Introdução

Sob o ponto de vista da Teoria de Superpopulações, há uma variável aleatória associada a cada elemento da população e o valor observado da característica de interesse é visto como uma realização dessa variável aleatória.

Nesta abordagem, a aleatorização está centrada em um modelo probabilístico, que gera a população em estudo. É suposto que o estatístico tenha alguma informação suplementar a respeito da população (variáveis auxiliares), que esteja de alguma forma relacionada com a variável de interesse, que, como já explicitado anteriormente, será representada por $Y_k(t)$.

A variável $X_{kp}(t)$ denotará a p -ésima variável auxiliar associada ao k -ésimo elemento da população $P_1(t)$, no instante t .

Neste capítulo, $P_2(T)$ será considerada contínua, $P_2(T) = \{t: 0 \leq t \leq T\}$, e o tipo de dados estará restrito ao painel completo, discreto ou contínuo.

Para a obtenção de informações mais detalhadas sobre a idéia conceitual de superpopulação ver Cassel/Särndal/Wretman (1977), Bolfarine/Rodrigues (1984) e Elian (1984).

3.2. Paine! Discreto

Suponha que n elementos da população $P_1(t)$ foram amostrados em um determinado instante e pesquisados para r instantes de $P_2(T)$. O conjunto de dados pode ser representado por:

$$\begin{array}{ccccccc} y_1(1) & y_2(1) & \dots\dots\dots & y_n(1) \\ y_1(2) & y_2(2) & \dots\dots\dots & y_n(2) \\ \vdots & \vdots & & \vdots \\ y_1(r) & y_2(r) & \dots\dots\dots & y_n(r) \end{array}$$

Considerando somente os r instantes de $P_2(T)$ que foram amostrados, o parâmetro populacional pode ser representado por

$$\tilde{Y}' = [\tilde{Y}'_1, \tilde{Y}'_2, \dots, \tilde{Y}'_N] ,$$

onde

$$\tilde{Y}'_k = [Y_k(1), Y_k(2), \dots, Y_k(r)] ,$$

que, após a seleção e observação da amostra, será parcialmente conhecido.

A matriz de variáveis auxiliares, totalmente conhecida, será dada por:

$$\underline{\tilde{x}} = \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_N \end{pmatrix} \quad \text{onde} \quad \underline{\tilde{x}}_k = \begin{pmatrix} x_{k1}(1) & x_{k2}(1) & \dots & x_{kp}(1) \\ x_{k1}(2) & x_{k2}(2) & \dots & x_{kp}(2) \\ \vdots & \vdots & & \vdots \\ x_{k1}(r) & x_{k2}(r) & \dots & x_{kp}(r) \end{pmatrix}$$

O modelo probabilístico, aqui denominado ζ , que as -
socia Y com X_p será dado por:

$$(A) \quad \zeta: \underline{y} = \underline{\tilde{x}}\underline{\beta} + \underline{e} \quad ,$$

onde $\underline{\beta}$ é um vetor de ordem $(p \times 1)$, de constantes desconhecidas,
representado por

$$\underline{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}$$

e $\underline{e}$ é um vetor de ordem $(N \times 1)$, representado por

$$\begin{pmatrix} \tilde{e}_1 \\ \tilde{e}_2 \\ \vdots \\ \tilde{e}_N \end{pmatrix} \quad ,$$

onde

$$\underline{e}_k' = [e_k(1), e_k(2), \dots, e_k(r)] ,$$

cuja média é zero, ou seja, $E[\underline{e}_k] = \underline{0}$ e cuja matriz de covariâncias, desconhecida, é dada por

$$V[\underline{e}] = \underline{V} = \text{diag}(\underline{V}_1, \underline{V}_2, \dots, \underline{V}_N) ,$$

onde $\underline{V}_k$ é a matriz de covariâncias de $\underline{e}_k$.

A suposição, feita acima, de que $\underline{V}$ é uma matriz bloco-diagonal, significa que as variáveis aleatórias correspondentes aos N elementos de $P_1(t)$ são, duas a duas, não correlacionadas.

Observações:

1. Se, por exemplo, $P_1(t)$ tivesse sido amostrada em um único instante, que poderia ser denotado pelo instante "1", as matrizes acima ficariam reduzidas a:

$$\underline{Y} = [Y_1(1), Y_2(1), \dots, Y_N(1)]$$

$$\underline{X} = \begin{pmatrix} x_{11}(1) & x_{12}(1) & \dots & x_{1p}(1) \\ x_{21}(1) & x_{22}(1) & \dots & x_{2p}(1) \\ \vdots & \vdots & & \vdots \\ x_{N1}(1) & x_{N2}(1) & \dots & x_{Np}(1) \end{pmatrix}$$

$$\tilde{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} \quad \tilde{e}' = [e_1(1), e_2(1), \dots, e_N(1)]$$

2. O modelo ζ é uma situação particular do modelo de regressão em N grupos considerado por Pfeiffermann (1984), Lindley/Smith (1972) e Fortney/Miller (1987).

Dada uma amostra e o modelo ζ introduzido em (A), pode-se, sem perda de generalidade, reordenar os vetores e matrizes em questão, de modo que os primeiros elementos sejam aqueles que pertencem à amostra. Então $\tilde{y}' = [\tilde{y}'_S, \tilde{y}'_{\bar{S}}]$, onde $\tilde{y}'_S$ é o vetor das componentes de $\tilde{y}$ que corresponde aos elementos da amostra e $\tilde{y}'_{\bar{S}}$ é o vetor referente aos elementos não amostrados, isto é, $\tilde{y}'_S = [\tilde{y}'_1, \tilde{y}'_2, \dots, \tilde{y}'_n]$ e $\tilde{y}'_{\bar{S}} = [\tilde{y}'_{n+1}, \dots, \tilde{y}'_N]$.

Analogamente, pode-se escrever:

$$\tilde{X} = \begin{pmatrix} X_S \\ \\ \\ X_{\bar{S}} \end{pmatrix}, \text{ onde } X_S = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ \vdots \\ \vdots \\ X_n \end{pmatrix}, \quad X_{\bar{S}} = \begin{pmatrix} X_{n+1} \\ X_{n+2} \\ \vdots \\ \vdots \\ \vdots \\ X_N \end{pmatrix},$$

$$\tilde{e}' = [\tilde{e}'_S, \tilde{e}'_{\bar{S}}] \quad \text{e} \quad V[\tilde{e}] = \text{diag}(V_S, V_{\bar{S}}),$$

onde

$$V_S = \text{diag}(V_1, V_2, \dots, V_n)$$

Como na teoria clássica, exposta anteriormente, o problema seria prever

$$\theta = \int_0^T c(t) \theta(t) dF(t) ,$$

que se resume em prever $\theta(t)$, uma vez que $c(t)$ e $F(t)$ são funções conhecidas.

Algumas definições, tiradas da "Teoria da Previsão em Populações Finitas" (Bolfarine/Rodrigues - 1984) serão dadas a seguir.

Definição 3.1 - O previsor $\hat{\theta}$ é dito ser ζ -não viciado (não viciado sob o modelo ζ) para θ , se $E_{\zeta}[\hat{\theta} - \theta] = 0$.

Definição 3.2 - O previsor $\hat{\theta}^*$ é dito ser ζ -ótimo (ótimo sob o modelo ζ) para θ se e somente se é ζ -não viciado para θ e, para qualquer outro previsor ζ -não viciado para θ , $\hat{\theta}$, vale o seguinte:

$$E_{\zeta}[\hat{\theta}^* - \theta]^2 \leq E_{\zeta}[\hat{\theta} - \theta]^2 .$$

Este trabalho se restringirá somente a funções paramétricas lineares, ou seja, $\theta(t)$ é uma função linear de $Y(t)$, isto é, $\theta(t) = \underline{\ell}_N' Y(t)$, onde $\underline{\ell}_N'$ é um vetor de ordem $N \times 1$.

Teorema 3.1 - Sob o modelo ζ , supondo que a função distribuição $F(t)$ é absolutamente contínua em $P_2(T)$, o previsor ζ -ótimo para θ é dado por:

$$\hat{\theta}^* = \int_0^T c(t) \mathbb{L}'_{\tilde{N}} X(t) \hat{\beta}_{\tilde{N}} dF(t), \quad (3.1)$$

onde

$$\hat{\beta}_{\tilde{N}} = \left[\sum_{k \in S_{\tilde{N}}} X_k' V_k^{-1} X_k \right]^{-1} \sum_{k \in S_{\tilde{N}}} X_k' V_k^{-1} Y_k \quad (3.2)$$

Prova: Supondo que vale trocar a esperança sob o sinal de integração, tem-se que:

$$\begin{aligned} \text{i)} \quad E_{\zeta} [\hat{\theta} - \theta] &= 0 \iff E_{\zeta} [\hat{\theta}] = \int_0^T c(t) \mathbb{L}'_{\tilde{N}} X(t) \beta_{\tilde{N}} dF(t) = \\ &= \mathbb{L}'_{\tilde{N}} \left(\int_0^T c(t) X(t) dF(t) \right) \beta_{\tilde{N}}, \end{aligned}$$

onde

$$\int_0^T c(t) X(t) dF(t)$$

é uma matriz de integrais. (ver resultado 3 do apêndice).

ii) Se $E_{\zeta} [\hat{\theta} - \theta] = 0$ e como $F(t)$ é absolutamente contínua em $P_2(T)$, tem-se que

$$E_{\zeta} [\hat{\theta} - \theta]^2 = \text{Var}[\hat{\theta}] + \text{Var}[\theta].$$

Logo, encontrar o estimador ζ -ótimo para θ equivale a encontrar o estimador não viciado para

$$\mathbb{L}'_{\tilde{N}} \left(\int_0^T c(t) X(t) dF(t) \right) \beta_{\tilde{N}},$$

de variância mínima.

Usando o teorema III do apêndice (Gauss Markov Generalizado), tem-se que o estimador ótimo para

$$\underset{\sim}{\ell}'_{\underset{\sim}{N}} \left(\int_0^T \underset{\sim}{c}(t) \underset{\sim}{X}(t) dF(t) \right) \underset{\sim}{\beta}$$

é dado por

$$\underset{\sim}{\ell}'_{\underset{\sim}{N}} \left(\int_0^T \underset{\sim}{c}(t) \underset{\sim}{X}(t) dF(t) \right) \underset{\sim}{\hat{\beta}}, \quad \text{onde } \underset{\sim}{\hat{\beta}} = (\underset{\sim}{X}'_{\underset{\sim}{S}} \underset{\sim}{V}_{\underset{\sim}{S}}^{-1} \underset{\sim}{X}_{\underset{\sim}{S}})^{-1} \underset{\sim}{X}'_{\underset{\sim}{S}} \underset{\sim}{V}_{\underset{\sim}{S}}^{-1} \underset{\sim}{Y}_{\underset{\sim}{S}}.$$

Logo,

$$\hat{\theta}^* = \int_0^T \underset{\sim}{c}(t) \underset{\sim}{\ell}'_{\underset{\sim}{N}} \underset{\sim}{X}(t) \underset{\sim}{\hat{\beta}} dF(t).$$

Observe que,

$$\begin{aligned} (\underset{\sim}{X}'_{\underset{\sim}{S}} \underset{\sim}{V}_{\underset{\sim}{S}}^{-1} \underset{\sim}{X}_{\underset{\sim}{S}}) &= (\underset{\sim}{X}'_1, \underset{\sim}{X}'_2, \dots, \underset{\sim}{X}'_n) \text{diag}(\underset{\sim}{V}_1^{-1}, \underset{\sim}{V}_2^{-1}, \dots, \underset{\sim}{V}_n^{-1}) \begin{pmatrix} \underset{\sim}{X}_1 \\ \underset{\sim}{X}_2 \\ \vdots \\ \underset{\sim}{X}_n \end{pmatrix} = \\ &= (\underset{\sim}{X}'_1 \underset{\sim}{V}_1^{-1}, \underset{\sim}{X}'_2 \underset{\sim}{V}_2^{-1}, \dots, \underset{\sim}{X}'_n \underset{\sim}{V}_n^{-1}) \begin{pmatrix} \underset{\sim}{X}_1 \\ \underset{\sim}{X}_2 \\ \vdots \\ \underset{\sim}{X}_n \end{pmatrix} = \sum_{k \in S} \underset{\sim}{X}'_k \underset{\sim}{V}_k^{-1} \underset{\sim}{X}_k \end{aligned}$$

Analogamente,

$$\underset{\sim}{X}'_{\underset{\sim}{S}} \underset{\sim}{V}_{\underset{\sim}{S}}^{-1} \underset{\sim}{Y}_{\underset{\sim}{S}} = \sum_{k \in S} \underset{\sim}{X}'_k \underset{\sim}{V}_k^{-1} \underset{\sim}{Y}_k.$$

Então,

$$\underset{\sim}{\hat{\beta}} = \left[\sum_{k \in S} \underset{\sim}{X}'_k \underset{\sim}{V}_k^{-1} \underset{\sim}{X}_k \right]^{-1} \sum_{k \in S} \underset{\sim}{X}'_k \underset{\sim}{V}_k^{-1} \underset{\sim}{Y}_k$$

Observações:

1. O previsor definido pelo teorema 3.1 é uma versão para populações dinâmicas do previsor tipo projeção, introduzido por Särndal/Wright (1984) e discutido com detalhes por Bolfarine/Rodrigues (1987).
2. O previsor introduzido pelo teorema 3.1 é simples e intuitivo pelo seguinte fato:

$$\hat{\theta}^* = \int_0^T c(t) \frac{1}{N} \hat{Y}(t) dF(t) \quad , \quad \text{onde} \quad \hat{Y}(t) = \frac{1}{N} \hat{X}(t) \hat{\beta} \quad .$$

Previsores simples e intuitivos com propriedades ótimas são de grande interesse, sem problemas práticos (ver Särndal/Wright - 1984).

3.3. Aplicação I - Caso Discreto

Para os exemplos, serão utilizadas a média populacional no instante t , $\theta(t) = \frac{1}{N} \frac{1}{N} Y(t)$ e a distribuição uniforme no intervalo $[0, T]$. Neste caso, θ será dado por

$$\theta = \frac{1}{NT} \int_0^T c(t) \frac{1}{N} Y(t) dt.$$

O modelo probabilístico considerado será o de locação, $Y(t) = X(t)\beta + e(t)$, onde $X(t) = \frac{1}{N}$, $\forall t \in P_2(T)$. Quanto à matriz de covariâncias, três casos serão examinados:

$$1) \quad \underline{V}_k = \sigma^2 \underline{I}$$

A variância é constante para os r instantes e não existe correlação entre informações correspondentes a um mesmo elemento k , para instantes diferentes.

Então, $\hat{\beta}$ será dado por:

$$\hat{\beta} = \left[\sum_{k \in S} \underline{1}'_r (\sigma^2 \underline{I}_r)^{-1} \underline{1}_r \right]^{-1} \sum_{k \in S} \underline{1}'_r (\sigma^2 \underline{I}_r)^{-1} \underline{y}_k$$

Mas,

$$\underline{1}'_r (\sigma^2 \underline{I}_r)^{-1} \underline{1}_r = \frac{r}{\sigma^2}$$

e

$$\underline{1}'_r (\sigma^2 \underline{I}_r)^{-1} \underline{y}_k = \frac{1}{\sigma^2} \sum_{i=1}^r y_k(i)$$

$$\hat{\beta} = \frac{\sigma^2}{nr} \frac{1}{\sigma^2} \sum_{k \in S} \sum_{i=1}^r y_k(i) = \frac{1}{nr} \sum_{k \in S} \sum_{i=1}^r y_k(i)$$

$$2) \quad \underline{V}_k = \sigma^2 \begin{pmatrix} 1 & \rho & \rho \dots \dots \rho \\ \rho & 1 & \rho \dots \dots \rho \\ \vdots & \vdots & \vdots \\ \rho & \rho & \rho \dots \dots 1 \end{pmatrix}$$

A variância é constante para os r instantes e existe correlação entre informações de um mesmo elemento, para instantes diferentes, também constante.

Neste caso, o cálculo de $\hat{\beta}$ seria: $\underline{V}_k^{-1} = \frac{1}{\sigma^2} \underline{W}_k^{-1}$, onde $\underline{W}_k^{-1}$ é obtida através do resultado 4 do apêndice, ou seja, é uma matriz formada pelos elementos

$$a_{ij} = \frac{1+(r-2)\rho}{(1+(r-1)\rho)(1-\rho)} \quad , \text{ para } i = j$$

e

$$a_{ij} = \frac{-\rho}{(1+(r-1)\rho)(1-\rho)} \quad , \text{ para } i \neq j$$

onde a_{ij} indica o elemento da i -ésima linha e j -ésima coluna.

$$V_{\sim k}^{-1} = \frac{1}{\sigma^2 (1+(r-1)\rho)(1-\rho)} \begin{pmatrix} 1+(r-2)\rho & -\rho & -\rho \\ -\rho & 1+(r-2)\rho & -\rho \\ \vdots & \vdots & \vdots \\ -\rho & -\rho & 1+(r-2)\rho \end{pmatrix}$$

$$X'_{\sim k} V_{\sim k}^{-1} X_{\sim k} = \frac{1+(r-2)\rho - (r-1)\rho}{\sigma^2 (1+(r-1)\rho)(1-\rho)} \mathbf{1}'_{\sim r} \mathbf{1}_{\sim r} = \frac{r}{\sigma^2 (1+(r-1)\rho)}$$

$$X'_{\sim k} V_{\sim k}^{-1} Y_{\sim k} = \frac{1}{\sigma^2 (1+(r-1)\rho)} \sum_{i=1}^r Y_k(i)$$

$$\hat{\beta} = \frac{\sigma^2 (1+(r-1)\rho)}{r} \frac{1}{\sigma^2 (1+(r-1)\rho)} \sum_{k \in S} \sum_{i=1}^r Y_k(i) =$$

$$= \frac{1}{nr} \sum_{k \in S} \sum_{i=1}^r Y_k(i) \quad .$$

$$3) \quad V_{\sim k} = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^{r-1} \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^{r-2} \\ \rho^2 & \rho & 1 & \rho & \dots & \rho^{r-3} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \rho^{r-1} & \rho^{r-2} & \rho^{r-3} & \dots & & 1 \end{pmatrix}$$

A variância é constante para os r instantes e a correlação entre informações correspondentes a um mesmo elemento, para dois instantes diferentes, vai diminuindo, segundo uma progressão geométrica, à medida que o intervalo de tempo cresce. A correlação entre $Y_k(t)$ e $Y_k(t+j)$ é dada por ρ^j , $t=1,2,\dots,r-1$ e $j=1,2,\dots,r-t$

$$V_k^{-1} = \frac{1}{\sigma^2(1-\rho^2)} \begin{pmatrix} 1 & -\rho & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix}$$

$$\tilde{X}_k' V_k^{-1} \tilde{X}_k = \frac{1}{\sigma^2(1-\rho^2)} (1-\rho, (1-\rho)^2, (1-\rho)^2, \dots, (1-\rho)^2, 1-\rho) \underline{1}_r =$$

$$= \frac{1}{\sigma^2(1+\rho)} (1, 1-\rho, 1-\rho, \dots, 1-\rho, 1) \underline{1}_r =$$

$$= \frac{1}{\sigma^2(1+\rho)} (2+(r-2)(1-\rho)) = \frac{r(1-\rho)+2\rho}{\sigma^2(1+\rho)}$$

$$\tilde{X}_k' V_k^{-1} \tilde{Y}_k = \frac{1}{\sigma^2(1+\rho)} (1, 1-\rho, 1-\rho, \dots, 1-\rho, 1) \tilde{Y}_k =$$

$$= \frac{1}{\sigma^2(1+\rho)} (Y_k(1) + Y_k(r) + (1-\rho) \sum_{i=2}^{r-1} Y_k(i))$$

$$\hat{\beta} = \frac{\sigma^2(1+\rho)}{n[r(1-\rho)+2\rho]} \frac{1}{\sigma^2(1+\rho)} \sum_{k \in S} (Y_k(1) + Y_k(r) + (1-\rho) \sum_{i=2}^{r-1} Y_k(i)) =$$

$$= \frac{1}{n[r(1-\rho)+2\rho]} \sum_{k \in S} (Y_k(1) + Y_k(r) + (1-\rho) \sum_{i=2}^{r-1} Y_k(i)) .$$

Nos dois primeiros casos, $\hat{\beta}$ é a média aritmética simples de todos os dados. Já no terceiro caso, $\hat{\beta}$ é a média aritmética ponderada, sendo que os dados correspondentes aos instantes 1 e r tem peso 1 e os demais tem peso $(1-\rho)$.

$$\text{Então, } \theta = \frac{1}{TN} \int_0^T c(t) \mathbf{1}'_{\tilde{N}} \tilde{Y}(t) dt .$$

Se $c(t) = c$ (constante), $\theta = \frac{c}{TN} \int_0^T \mathbf{1}'_{\tilde{N}} \tilde{Y}(t) dt$

e $\hat{\theta}^* = \frac{c}{TN} \int_0^T \mathbf{1}'_{\tilde{N}} \mathbf{1}_{\tilde{N}} \hat{\beta} dt = \frac{cNT}{TN} \hat{\beta} = c\hat{\beta} .$

Quatro situações serão consideradas e os previsores encontrados serão iguais para os dois primeiros tipos de matriz de covariâncias, acima mencionadas.

Então, para $\forall t \in P_2(T)$, $c(t) = c$:

a) $c = 1$, θ seria a média das médias, com relação ao intervalo $[0, T]$.

para i) e ii) $\hat{\theta}^* = \frac{1}{nr} \sum_{k \in S} \sum_{i=1}^r Y_k(i)$

para iii) $\hat{\theta}^* = \frac{1}{n[r(1-\rho)+2\rho]} \sum_{k \in S} (Y_k(1) + Y_k(r) + (1-\rho) \sum_{i=2}^{r-1} Y_k(i))$

b) $c = T$, o significado de θ seria o de acompanhar o comportamento da média ao longo do tempo

para i) e ii) $\hat{\theta}^* = \frac{T}{nr} \sum_{k \in S} \sum_{i=1}^r Y_k(i)$

para iii) $\hat{\theta}^* = \frac{T}{n[r(1-\rho)+2\rho]} \sum_{k \in S} (Y_k(1) + Y_k(r) + (1-\rho) \sum_{i=2}^{r-1} Y_k(i))$

c) $c = N$, θ seria a média dos totais com relação ao intervalo $[0, T]$.

para i) e ii) $\hat{\theta}^* = \frac{N}{nr} \sum_{k \in S} \sum_{i=1}^r Y_k(i)$

para iii) $\hat{\theta}^* = \frac{N}{n[r(1-\rho)+2\rho]} \sum_{k \in S} (Y_k(1) + Y_k(r) + (1-\rho) \sum_{i=2}^{r-1} Y_k(i))$

d) $c = TN$, o significado de θ seria o de acompanhar o comportamento do total ao longo do tempo

para i) e ii) $\hat{\theta}^* = \frac{TN}{nr} \sum_{k \in S} \sum_{i=1}^r Y_k(i)$

para iii) $\hat{\theta}^* = \frac{TN}{n[r(1-\rho)+2\rho]} \sum_{k \in S} (Y_k(1) + Y_k(r) + (1-\rho) \sum_{i=2}^{r-1} Y_k(i))$

Quando $N = 1$, $\theta = \int_0^T c(t) Y(t) dF(t)$. Como exemplo,

pode-se citar o estudo do comportamento da temperatura $Y(t)$ de uma região, ao longo do tempo. Para maiores detalhes, ver Bartlett (1986). Neste caso particular, para $F(t)$ uniforme

e $c=1$, o previsor ζ -ótimo para $\theta = \frac{1}{T} \int_0^T Y(t) dt$ nos casos i) e ii) seria

$$\hat{\theta}^* = \frac{1}{r} \sum_{i=1}^r Y(i) \quad , \text{ onde } \{1, 2, \dots, r\}$$

seria o conjunto dos instantes selecionados na população contínua $P_2(T)$.

3.4. Painel Contínuo

Nesta seção, será feita a suposição de que n elementos da população $P_1(t)$ foram amostrados no instante "0" e que, cada um deles foi observado durante o intervalo $[0, T]$.

O conjunto de dados será o conjunto de n trajetórias nesse intervalo, ou seja,

No instante t , o parâmetro populacional será dado por $\underline{y}(t) = [Y_1(t), Y_2(t), \dots, Y_N(t)]$, a matriz de variáveis auxiliares por

$$\underline{\tilde{x}}(t) = \begin{pmatrix} \underline{\tilde{x}}_1(t) \\ \underline{\tilde{x}}_2(t) \\ \vdots \\ \underline{\tilde{x}}_N(t) \end{pmatrix}, \text{ onde } \underline{\tilde{x}}_k(t) = (x_{k1}(t), x_{k2}(t), \dots, x_{kp}(t)),$$

o vetor de erros por $\underline{e}(t) = [e_1(t), e_2(t), \dots, e_N(t)]$ e o modelo ζ por $\underline{y}(t) = \underline{\tilde{x}}(t)\underline{\beta} + \underline{e}(t)$, onde, como no caso discreto, $E[\underline{e}(t)] = \underline{0}$ e $\underline{V}(t)$, a matriz de covariâncias de $\underline{e}(t)$, é diagonal, ou seja, não existe correlação entre informações referentes a elementos diferentes, no mesmo instante. A função $\underline{\tilde{x}}(t)\underline{\beta}$ é denominada por Bartlett (1986) como função de tendência.

Serão considerados somente os casos em que $\theta(t)$ é uma combinação linear de $Y_k(t)$, ou seja,

$$\theta(t) = \underline{\ell}'_{N\sim} \underline{Y}(t) \quad \text{e} \quad \theta = \int_0^T c(t) \underline{\ell}'_{N\sim} \underline{Y}(t) dF(t).$$

Reordenando os vetores e matrizes em questão, θ pode ser escrito da seguinte forma:

$$\theta = \int_0^T c(t) \underline{\ell}'_{n\sim s} \underline{Y}_s(t) dF(t) + \int_0^T c(t) \underline{\ell}'_{N-n\sim \tilde{s}} \underline{Y}_{\tilde{s}}(t) dF(t).$$

Após a seleção e observação da amostra, a primeira parte será conhecida e então, o problema passa a ser prever a segunda parcela, usando os dados da amostra.

Entretanto, pode-se escrever o modelo da seguinte maneira:

$$\underline{Y}^* = \underline{X}^* \underline{\beta} + \underline{e}^*$$

onde

$$\underline{Y}^* = \begin{pmatrix} \int_0^T c(t) Y_1(t) dF(t) \\ \int_0^T c(t) Y_2(t) dF(t) \\ \vdots \\ \int_0^T c(t) Y_N(t) dF(t) \end{pmatrix},$$

que será representado por $\underline{Y}^* = \int_0^T c(t) \underline{Y}(t) dF(t)$. Analogamente,

$$\underline{X}^* = \int_0^T c(t) \underline{X}(t) dF(t) \quad \text{e} \quad \underline{e}^* = \int_0^T c(t) \underline{e}(t) dF(t).$$

$$E[\underline{e}^*] = E\left[\int_0^T c(t) \underline{e}(t) dF(t)\right] = \int_0^T c(t) E[\underline{e}(t)] dF(t) = \underline{0}$$

e

$$\text{Var}[\underline{e}_k^*] = E\left[\int_0^T c(t) e_k(t) dF(t)\right]^2$$

A matriz de covariâncias de $\underline{e}^*$ será denotada por $\underline{V}^*$ e

$$\underline{V}^* = \text{diag}(V_1^*, V_2^*, \dots, V_N^*) \quad \text{onde} \quad V_k^* = \text{Var}[e_k^*]$$

Utilizando-se o teorema IV do apêndice (Gauss Markov para Populações Finitas), tem-se que o previsor linear ζ -ótimo para θ é dado por:

$$\begin{aligned}\hat{\theta}^* &= \int_0^T c(t) \tilde{y}_S^*(t) dF(t) + \int_0^T c(t) \tilde{x}_{N-n}^*(t) dF(t) \hat{\beta} \\ &= \int_0^T c(t) \tilde{y}_S^*(t) dF(t) + \int_0^T c(t) \tilde{x}_{N-n}^*(t) dF(t) \hat{\beta}\end{aligned}\quad (3.3)$$

onde

$$\hat{\beta} = (\tilde{x}_S^{*'} \tilde{V}_S^{*-1} \tilde{x}_S^*)^{-1} \tilde{x}_S^{*'} \tilde{V}_S^{*-1} \tilde{y}_S^* \quad (3.4)$$

3.5. Aplicação II - Caso Contínuo

Serão considerados, novamente, $\theta(t) = \frac{1}{N} \frac{1}{N} y(t)$ e $F(t)$ a distribuição uniforme em $[0, T]$. Quanto ao modelo, serão examinados dois casos: modelo de locação e modelo de regressão linear simples passando pela origem.

i) Modelo de Locação

$$y(t) = x(t)\beta + e(t), \quad x(t) = \frac{1}{N}$$

$$\begin{aligned}\tilde{x}_S^{*'} \tilde{V}_S^{*-1} \tilde{x}_S^* &= \int_0^T c(t) \frac{1}{N} dF(t) \text{diag} \left(\frac{1}{V_1^*}, \frac{1}{V_2^*}, \dots, \frac{1}{V_n^*} \right) \int_0^T c(t) \frac{1}{N} dF(t) = \\ &= \sum_{k \in S} \left(\int_0^T c(t) dF(t) \right)^2 \frac{1}{V_k^*} = \left(\int_0^T c(t) dF(t) \right)^2 \sum_{k \in S} \frac{1}{V_k^*}\end{aligned}$$

$$\begin{aligned} \mathbf{X}_S^{*'} \mathbf{Y}_S^{*-1} \mathbf{Y}_S^{*} &= \int_0^T \mathbf{c}(t) \mathbf{I}_n' dF(t) \text{diag} \left(\frac{1}{V_1^*}, \frac{1}{V_2^*}, \dots, \frac{1}{V_n^*} \right) \int_0^T \mathbf{c}(t) \mathbf{Y}_S(t) dF(t) = \\ &= \sum_{k \in S} \int_0^T \mathbf{c}(t) dF(t) \frac{1}{V_k^*} \int_0^T \mathbf{c}(t) Y_k(t) dF(t) = \\ &= \int_0^T \mathbf{c}(t) dF(t) \sum_{k \in S} \frac{1}{V_k^*} \int_0^T \mathbf{c}(t) Y_k(t) dF(t). \end{aligned}$$

$$\begin{aligned} \hat{\beta} &= \frac{1}{\left(\int_0^T \mathbf{c}(t) dF(t) \right)^2 \sum_{k \in S} \frac{1}{V_k^*}} \int_0^T \mathbf{c}(t) dF(t) \sum_{k \in S} \frac{1}{V_k^*} \int_0^T \mathbf{c}(t) Y_k(t) dF(t) = \\ &= \frac{1}{\int_0^T \mathbf{c}(t) dF(t) \sum_{k \in S} \frac{1}{V_k^*}} \sum_{k \in S} \frac{1}{V_k^*} \int_0^T \mathbf{c}(t) Y_k(t) dF(t). \end{aligned}$$

Supondo que a matriz de covariâncias de $\underline{e}^*$ é $\underline{V}^* = \sigma^2 \underline{I}$ e substituindo a distribuição $F(t)$,

$$\begin{aligned} \hat{\beta} &= \frac{1}{\int_0^T \mathbf{c}(t) \frac{1}{T} dt \cdot n} \sum_{k \in S} \int_0^T \mathbf{c}(t) Y_k(t) \frac{1}{T} dt = \\ &= \frac{1}{n \int_0^T \mathbf{c}(t) dt} \sum_{k \in S} \int_0^T \mathbf{c}(t) Y_k(t) dt. \end{aligned}$$

Então,

$$\begin{aligned}
 \hat{\theta}^* &= \frac{1}{N-n} \int_0^T c(t) Y_s(t) \frac{1}{T} dt + \frac{1}{N-n} \int_0^T c(t) \frac{1}{N-nT} dt \frac{1}{n \int_0^T c(t) dt} \sum_{k \in S} \int_0^T c(t) Y_k(t) dt = \\
 &= \frac{1}{TN} \sum_{k \in S} \int_0^T c(t) Y_k(t) dt + \frac{1}{NT} \sum_{k \in S} \int_0^T c(t) dt \frac{1}{n \int_0^T c(t) dt} \sum_{k \in S} \int_0^T c(t) Y_k(t) dt = \\
 &= \frac{1}{TN} \left(1 + \sum_{k \in S} \int_0^T c(t) dt \frac{1}{n \int_0^T c(t) dt} \right) \sum_{k \in S} \int_0^T c(t) Y_k(t) dt .
 \end{aligned}$$

Se $c(t)$ é constante para $\forall t \in P_2(T)$, ou seja, $c(t) = c$, então:

$$\hat{\beta} = \frac{1}{nT} \sum_{k \in S} \int_0^T Y_k(t) dt$$

e

$$\begin{aligned}
 \hat{\theta}^* &= \frac{1}{TN} \left(1 + \frac{(N-n) cT}{ncT} \right) c \sum_{k \in S} \int_0^T Y_k(t) dt = \\
 &= \frac{n+N-n}{TNn} c \sum_{k \in S} \int_0^T Y_k(t) dt = \frac{c}{nT} \sum_{k \in S} \int_0^T Y_k(t) dt .
 \end{aligned}$$

Outra vez, serão examinadas quatro situações diferentes, de acordo com a função $c(t) = c$.

$$a) \ c = 1, \quad \hat{\theta}^* = \frac{1}{nT} \sum_{k \in S} \int_0^T Y_k(t) dt$$

$$b) \ c = T, \quad \hat{\theta}^* = \frac{1}{n} \sum_{k \in S} \int_0^T Y_k(t) dt$$

$$c) \ c = N, \quad \hat{\theta}^* = \frac{N}{nT} \sum_{k \in S} \int_0^T Y_k(t) dt$$

$$d) \ c = TN, \quad \hat{\theta}^* = \frac{N}{n} \sum_{k \in S} \int_0^T Y_k(t) dt \text{ (conhecido na literatura como}$$

estimador expansão - Herson/Royall - 1973).

ii) Modelo de Regressão Linear Simples, passando pela origem.

$$\underline{Y}(t) = \underline{X}(t) \underline{\beta} + \underline{e}(t)$$

$$\underline{X}'(t) = (x_1, x_2, \dots, x_N)$$

$$\begin{aligned} \underline{X}'_S * \underline{V}_S^{*-1} \underline{X}_S^* &= \int_0^T c(t) (x_1, x_2, \dots, x_n) dF(t) \text{diag} \frac{1}{V_1^*}, \frac{1}{V_2^*}, \dots, \frac{1}{V_n^*} \int_0^T c(t) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} dF(t) = \\ &= \sum_{k \in S} \left(\int_0^T c(t) x_k dF(t) \right)^2 \frac{1}{V_k^*} = \left(\int_0^T c(t) dF(t) \right)^2 \sum_{k \in S} \frac{x_k^2}{V_k^*} \end{aligned}$$

$$\begin{aligned} \underline{X}'_S * \underline{V}_S^{*-1} \underline{Y}_S^* &= \sum_{k \in S} \int_0^T c(t) x_k dF(t) \frac{1}{V_k^*} \int_0^T c(t) Y_k(t) dF(t) = \\ &= \sum_{k \in S} \int_0^T c(t) dF(t) \frac{x_k}{V_k^*} \int_0^T c(t) Y_k(t) dF(t) = \\ &= \int_0^T c(t) dF(t) \sum_{k \in S} \frac{x_k}{V_k^*} \int_0^T c(t) Y_k(t) dF(t). \end{aligned}$$

$$\hat{\beta} = \frac{1}{\int_0^T c(t) dF(t) \sum_{k \in S} \frac{x_k^2}{V_k^*}} \sum_{k \in S} \frac{x_k}{V_k^*} \int_0^T c(t) Y_k(t) dF(t)$$

Supondo $\underline{V}^* = \sigma^2 \text{diag}(x_1, x_2, \dots, x_N)$ e substituindo a distribuição $F(t)$,

$$\hat{\beta} = \frac{1}{\int_0^T c(t) \frac{1}{T} dt \sum_{k \in S} x_k} \sum_{k \in S} \int_0^T c(t) y_k(t) \frac{1}{T} dt =$$

$$= \frac{1}{\int_0^T c(t) dt \sum_{k \in S} x_k} \sum_{k \in S} \int_0^T c(t) y_k(t) dt$$

e

$$\hat{\theta}^* = \frac{1}{N} l'_n \int_0^T c(t) y_s(t) \frac{1}{T} dt + \frac{1}{N} l'_{N-n} \int_0^T c(t) x_{\tilde{s}}(t) \frac{1}{T} dt \frac{1}{\int_0^T c(t) dt \sum_{k \in S} x_k} \sum_{k \in S} \int_0^T c(t) y_k(t) dt =$$

$$= \frac{1}{TN} \sum_{k \in S} \int_0^T c(t) y_k(t) dt + \frac{1}{TN} \sum_{k \notin S} \int_0^T c(t) x_k dt \frac{1}{\int_0^T c(t) dt \sum_{k \in S} x_k} \sum_{k \in S} \int_0^T c(t) y_k(t) dt =$$

$$= \frac{1}{TN} \left(1 + \sum_{k \notin S} \int_0^T c(t) x_k dt \frac{1}{\int_0^T c(t) dt \sum_{k \in S} x_k} \right) \sum_{k \in S} \int_0^T c(t) y_k(t) dt .$$

Se $c(t) = c$ (constante),

$$\hat{\beta} = \frac{1}{T \sum_{k \in S} x_k} \sum_{k \in S} \int_0^T y_k(t) dt$$

e

$$\hat{\theta}^* = \frac{1}{TN} \left(1 + c T \sum_{k \notin S} x_k \frac{1}{c T \sum_{k \in S} x_k} \right) c \sum_{k \in S} \int_0^T y_k(t) dt =$$

$$= \frac{\sum_{k \in S} x_k + \sum_{k \notin S} x_k}{TN \sum_{k \in S} x_k} c \sum_{k \in S} \int_0^T y_k(t) dt = \frac{\sum_{k=1}^N x_k}{TN \sum_{k \in S} x_k} c \sum_{k \in S} \int_0^T y_k(t) dt .$$

$$a) \text{ Se } c = 1, \hat{\theta}^* = \frac{1}{TN} \frac{\sum_{k=1}^N x_k}{\sum_{k \in S} x_k} \sum_{k \in S} \int_0^T y_k(t) dt$$

$$b) \text{ Se } c = T, \hat{\theta}^* = \frac{1}{N} \frac{\sum_{k=1}^N x_k}{\sum_{k \in S} x_k} \sum_{k \in S} \int_0^T y_k(t) dt$$

$$c) \text{ Se } c = N, \hat{\theta}^* = \frac{1}{T} \frac{\sum_{k=1}^N x_k}{\sum_{k \in S} x_k} \sum_{k \in S} \int_0^T y_k(t) dt$$

$$d) \text{ Se } c = TN, \hat{\theta}^* = \frac{\sum_{k=1}^N x_k}{\sum_{k \in S} x_k} \sum_{k \in S} \int_0^T y_k(t) dt \quad (\text{conhecido na li-}$$

teratura como previsor razão - Cochran (1963) - Rao (1973)).

oOo

CAPÍTULO 4

OBTENÇÃO DE ESTIMADORES ÓTIMOS ATRAVÉS DA TEORIA DE SUPERPOPULAÇÃO

4.1. Introdução

O objetivo deste capítulo é obter estimadores ótimos para $E_{\zeta}(\theta)$, através da Teoria de Superpopulações, para vários esquemas amostrais adotados. Como na teoria clássica, serão considerados estimadores do tipo convexo, isto é,

$$\hat{\theta}(t) = q \cdot \hat{\theta}(0) + (1-q) \hat{\theta}(d(t)),$$

para $0 \leq q \leq 1$, conforme notação do capítulo 2.

O interesse em estimar $E_{\zeta}(\theta)$ foi motivado por Firouzabadi (1975) e Tam (1986). Firouzabadi denotou por X a característica de interesse no instante um e por Y a mesma característica no instante dois. Supondo que (X, Y) tem distribuição bivariada com vetor de médias

$$\tilde{\theta} = \begin{pmatrix} E(X) \\ E(Y) \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$

e matriz de covariâncias

$$\tilde{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

ele apresentou estimadores para θ_1 , θ_2 e $\theta_2 - \theta_1$.

Critério para otimalidade

O critério para obtenção do estimador será o seguinte: fixado o estimador $\hat{\theta}(d(t))$, será encontrado o valor de "q" e o estimador $\hat{\theta}(p)$ que torna $\hat{\theta}$ ótimo para $E_{\zeta}(\theta)$.

Notações:

$$1. \quad u_t = \{\hat{\theta}(t) : E_{\zeta}[\hat{\theta}(p)] = E_{\zeta}[\theta(t)] \quad \text{e} \quad E_{\zeta}[\hat{\theta}(d(t))] = E_{\zeta}[\theta(t)]\}$$

u_t é a classe dos estimadores de $E_{\zeta}[\theta(t)]$ não viciados, cujos componentes $\hat{\theta}(p)$ e $\hat{\theta}(d(t))$ também são não viciados para $E_{\zeta}[\theta(t)]$.

Note que, se $\hat{\theta}(t) \in u_t$ então $\hat{\theta} = \int_0^T c(t) \hat{\theta}(t) dF(t)$ é não viciado para $E_{\zeta}(\theta)$.

$$2. \quad u_c = \{\hat{\theta} : \hat{\theta}(t) \in u_t\}$$

u_c é a classe dos estimadores de $E_{\zeta}(\theta)$, cujo componente $\hat{\theta}(t)$ pertence à classe u_t .

Definição 4.1 - O estimador $\hat{\theta}^*$ é ζ_c -ótimo para θ se e somente se satisfaz as duas condições abaixo:

- i) $\hat{\theta}^* \in u_c$
 ii) $\text{Var}_\zeta[\hat{\theta}^*] \leq \text{Var}_\zeta[\hat{\theta}]$, $\forall \hat{\theta} \in u_c$.

4.2. Procedimento para a obtenção de Estimadores Ótimos

Lema 4.1 - Suponha que exista um $\hat{\theta}$, tal que,

$$E_\zeta[\hat{\theta}(\mathcal{D}) - \theta(t)] = 0 \quad \text{e} \quad E_\zeta[\hat{\theta}(d(t)) - \theta(t)] = 0$$

- i) Se $\hat{\theta}_0$ é um estimador tal que $E_\zeta[\hat{\theta}_0] = 0$ e $\hat{\theta}_0 \neq 0$, então $\hat{\theta}^*$ é ζ_c -ótimo para $E_\zeta(\theta)$ se e somente se

$$q^* = \frac{\text{cov}_\zeta \left(\int_0^T c(t) \hat{\theta}(d(t)) dF(t), \hat{\theta}_0 \right)}{\text{cov}_\zeta \left(\int_0^T c(t) \hat{\theta}(d(t)) dF(t), \hat{\theta}_0 \right) - \text{cov}_\zeta \left(\int_0^T c(t) \hat{\theta}(\mathcal{D}) dF(t), \hat{\theta}_0 \right)} \quad (4.1)$$

- ii) $\text{Var}_\zeta[\hat{\theta}^*] = \text{Cov}_\zeta[\hat{\theta}^*, \hat{\theta}]$, onde $\hat{\theta} \in u_c$.

Prova. Se $E_\zeta[\hat{\theta}(\mathcal{D}) - \theta(t)] = 0$ e $E_\zeta[\hat{\theta}(d(t)) - \theta(t)] = 0$, não é difícil verificar que $E_\zeta[\hat{\theta}^* - \theta] = 0$. Segue do Teorema II do apêndice que:

$$\begin{aligned} \hat{\theta}^* \text{ é } \zeta_c\text{-ótimo para } E_\zeta[\theta] \\ \iff \text{Cov}_\zeta[\hat{\theta}^*, \hat{\theta}_0] = 0 &\iff E_\zeta[(\hat{\theta}^* - E_\zeta[\theta])(\hat{\theta}_0 - 0)] = 0 \\ \iff E_\zeta[\hat{\theta}^* \hat{\theta}_0 - E_\zeta[\theta] \hat{\theta}_0] = 0 &\iff E_\zeta[\hat{\theta}^* \hat{\theta}_0] - 0 = 0 \\ \iff E_\zeta[\hat{\theta}^* \hat{\theta}_0] = 0 &\iff E_\zeta[(q \cdot a + (1-q)b) \hat{\theta}_0] = 0 \quad , \end{aligned}$$

onde

$$a = \int_0^T c(t) \hat{\theta}(\mathcal{D}) dF(t) \quad e \quad b = \int_0^T c(t) \hat{\theta}(d(t)) dF(t) .$$

Logo,

$$E_{\zeta} [q \cdot a \hat{\theta}_0 + (1-q) b \cdot \hat{\theta}_0] = 0 \iff q \cdot E_{\zeta} [a \hat{\theta}_0] + (1-q) E_{\zeta} [b \hat{\theta}_0] = 0$$

$$\iff q \cdot \text{cov}_{\zeta} [a, \hat{\theta}_0] + (1-q) \text{cov}_{\zeta} [b, \hat{\theta}_0] = 0$$

$$\iff q = \frac{\text{cov}_{\zeta} [b, \hat{\theta}_0]}{\text{cov}_{\zeta} [b, \hat{\theta}_0] - \text{cov}_{\zeta} [a, \hat{\theta}_0]} .$$

O resultado ii) é uma aplicação direta do Teorema II do apêndice.

Teorema 4.1 - Se $\hat{\theta}^*(\mathcal{D})$ é o estimador ζ -ótimo para $E_{\zeta} [\theta(t)]$, então o estimador ζ_C -ótimo para $E_{\zeta} [\theta]$ é dado por:

$$\hat{\theta}^* = \int_0^T c(t) \hat{\theta}^*(\mathcal{D}) dF(t) \quad (4.2)$$

Prova: $\hat{\theta}^*(\mathcal{D})$ é ζ -ótimo para $E_{\zeta} [\theta(t)] \iff \text{cov}_{\zeta} [\hat{\theta}^*(\mathcal{D}), \hat{\theta}_0] = 0$.

Segundo o lema 4.1, o valor de "q" que corresponde ao estimador ζ_C -ótimo $\hat{\theta}^*$ é dado por

$$q^* = \frac{\text{cov}_{\zeta} [b, \hat{\theta}_0]}{\text{cov}_{\zeta} [b, \hat{\theta}_0] - \text{cov}_{\zeta} [a, \hat{\theta}_0]} .$$

Logo,

$$\begin{aligned} q^* &= \frac{\text{cov}_{\zeta} [b, \hat{\theta}_0]}{\text{cov}_{\zeta} [b, \hat{\theta}_0] - \text{cov}_{\zeta} \left[\int_0^T c(t) \hat{\theta}^*(\mathcal{D}) dF(t), \hat{\theta}_0 \right]} = \\ &= \frac{\text{cov}_{\zeta} [b, \hat{\theta}_0]}{\text{cov}_{\zeta} [b, \hat{\theta}_0] - E_{\zeta} \left[\left(\int_0^T c(t) \hat{\theta}^*(\mathcal{D}) dF(t) \right) \cdot \hat{\theta}_0 \right]} = \end{aligned}$$

$$\begin{aligned}
 &= \frac{\text{cov}_\zeta [b, \hat{\theta}_0]}{\text{cov}_\zeta [b, \hat{\theta}_0] - E_\zeta \left(\int_0^T c(t) \hat{\theta}^*(\mathcal{D}) \hat{\theta}_0 dF(t) \right)} = \\
 &= \frac{\text{cov}_\zeta [b, \hat{\theta}_0]}{\text{cov}_\zeta [b, \hat{\theta}_0] - \int_0^T c(t) E_\zeta [\hat{\theta}^*(\mathcal{D}) \hat{\theta}_0] dF(t)} = \\
 &= \frac{\text{cov}_\zeta [b, \hat{\theta}_0]}{\text{cov}_\zeta [b, \hat{\theta}_0] - \int_0^T c(t) \text{cov}_\zeta [\hat{\theta}^*(\mathcal{D}), \hat{\theta}_0] dF(t)} = \\
 &= \frac{\text{cov}_\zeta [b, \hat{\theta}_0]}{\text{cov}_\zeta [b, \hat{\theta}_0] - \int_0^T c(t) 0 \cdot dF(t)} = \frac{\text{cov}_\zeta [b, \hat{\theta}_0]}{\text{cov}_\zeta [b, \hat{\theta}_0]} = 1
 \end{aligned}$$

Logo, $\hat{\theta}^*$ é dado por $\hat{\theta}^* = \int_0^T c(t) \hat{\theta}^*(\mathcal{D}) dF(t)$.

Então, pelo Teorema 4.1, se o estimador $\hat{\theta}^*(\mathcal{D})$, ζ -ótimo para $E_\zeta [\theta(t)]$, for obtido, basta calcular $\int_0^T c(t) \hat{\theta}^*(\mathcal{D}) dF(t)$ para se obter $\hat{\theta}^*$. O resultado desse teorema é intuitivo, já que $\hat{\theta}^*(\mathcal{D})$ acumula toda a informação até o instante t e é ζ -ótimo para $E_\zeta [\theta(t)]$, $\forall t \in P_2(T)$.

O problema, portanto, consiste em encontrar $\hat{\theta}^*(\mathcal{D})$, o que pode ser feito através da técnica utilizada por Firouzabadi (1975), pela aplicação dos teoremas I e II do apêndice. É importante observar que, para N suficientemente grande, estimar $E_\zeta (\theta)$ é equivalente a estimar θ .

4.3. Aplicação I - Caso Discreto (duas ocasiões)

A seguir, nesta e nas próximas seções, serão vistas algumas aplicações dos procedimentos para obter estimadores ótimos, que foram introduzidos na seção anterior.

Em primeiro lugar, será examinado o caso em que os elementos são observados em duas ocasiões, 0 e 1, e a matriz de covariância do par $(Y_k(0), Y_k(1))$ é dada por

$$\Sigma = \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix},$$

conhecida.

Além disso, será feita a suposição de que não existe correlação entre distintas unidades de $P_1(t)$.

Alguns problemas de interesse serão relacionados a seguir. Supondo que $\theta(t)$ é a média populacional no instante t , $\theta(t) = \frac{1}{N} \sum_{i=1}^N Y_i(t)$, deseja-se estimar:

- a) $E_{\zeta}[\theta(0)] = E_{\zeta}[Y_k(0)]$, $k \in P_1(t)$
- b) $E_{\zeta}[\theta(1)] = E_{\zeta}[Y_k(1)]$, $k \in P_1(t)$
- c) $E_{\zeta}[\theta(1) - \theta(0)] = E_{\zeta}[Y_k(1)] - E_{\zeta}[Y_k(0)]$, $k \in P_1(t)$.

Esses três casos, aqui apresentados como exemplos, foram examinados por Firouzabadi (1975), que considerou dois tipos de dados: painel discreto incompleto e painel rotativo.

I - Painel discreto incompleto

O conjunto de dados será representado por:

$$y_1(0) \dots y_m(0) y_{m+1}(0) \dots y_n(0)$$

$$y_1(1) \dots y_m(1)$$

o que significa que m elementos responderam à pesquisa nos tempos 0 e 1 e u=n-m responderam somente no tempo 1.

A proporção dos que responderam duas vezes, sobre o tamanho da amostra, será denotada por λ .

$$\lambda = \frac{m}{n} \quad e \quad K = \frac{\sigma_1}{\sigma_0}.$$

A função F(t) será dada por

$$F(t) = \begin{cases} 0 & t < 0 \\ p_0 & 0 \leq t < 1 \\ p_0 + p_1 = 1 & t \geq 1 \end{cases}$$

a) Objetivo: adotando um painel discreto e um modelo ζ , o problema consiste em estimar $E_{\zeta}[\theta(0)]$. Neste caso especial,

$$c(t) = \begin{cases} 1/p_0 & , \quad p|t = 0 \\ 0 & , \quad cc \end{cases}$$

e

$$\theta = \int_0^T c(t) \theta(t) dF(t) = \frac{1}{p_0} \theta(0) p_0 = \theta(0). \quad (4.3)$$

Pelo Teorema 4.1, basta encontrar $\hat{\theta}^*(\mathcal{D})$, que por (4.3) será o próprio $\hat{\theta}^*$.

Para calcular $\hat{\theta}^*(\mathcal{D})$ basta considerar primeiramente o Teorema I do apêndice:

Seja $\tilde{P}$ a matriz de dados

$$\tilde{P} = \begin{pmatrix} y_1(0) & \dots & y_m(0) & y_{m+1}(0) & \dots & y_n(0) \\ y_1(1) & \dots & y_m(1) & & & \end{pmatrix}$$

Então, $\hat{\theta}^*$ terá a seguinte forma:

$$\begin{aligned} \hat{\theta}^* &= a'y_1(0) + \dots + a'y_m(0) + b'y_{m+1}(0) + \dots + b'y_n(0) + \\ &+ c'y_1(1) + \dots + c'y_m(1) = \\ &= a' \sum_{i=1}^m y_i(0) + b' \sum_{i=m+1}^n y_i(0) + c' \sum_{i=1}^m y_i(1). \end{aligned}$$

Usando
$$\bar{y}_m(0) = \frac{1}{m} \sum_{i=1}^m y_i(0)$$

$$\bar{y}_u(0) = \frac{1}{u} \sum_{i=m+1}^n y_i(0)$$

$$\bar{y}_m(1) = \frac{1}{m} \sum_{i=1}^m y_i(1) ,$$

tem-se que

$$\hat{\theta}^* = a\bar{y}_m(0) + b\bar{y}_u(0) + c\bar{y}_m(1) .$$

Para que $\hat{\theta}^*$ seja não viciado para $E_{\zeta}[\theta(0)]$, é suficiente escolher a, b e c tais que $a+b = 1$ e $c = 0$, ou seja,

$$\hat{\theta}^* = a\bar{Y}_m(0) + (1-a)\bar{Y}_u(0).$$

Seja $\hat{\theta}_0 = \bar{Y}_m(0) - \bar{Y}_u(0)$ um estimador não viciado, não trivial de zero. Logo, pelo Teorema II do apêndice, para que $\hat{\theta}^*$ seja ζ_c -ótimo para $E_\zeta[\theta(0)]$, basta que $\text{cov}_\zeta[\hat{\theta}^*, \hat{\theta}_0] = 0$. Portanto,

$$\begin{aligned} \text{cov}_\zeta[\hat{\theta}^*, \hat{\theta}_0] &= \text{cov}_\zeta[a\bar{Y}_m(0) + (1-a)\bar{Y}_u(0), \bar{Y}_m(0) - \bar{Y}_u(0)] = \\ &= a\frac{1}{m}\sigma_0^2 - (1-a)\frac{1}{u}\sigma_0^2 = \frac{au - m(1-a)}{mu}\sigma_0^2 = 0 \end{aligned}$$

ou $au - m(1-a) = 0$, $a = \frac{m}{u+m} = \frac{m}{n}$

logo
$$\begin{aligned} \hat{\theta}^* &= \frac{m}{n}\bar{Y}_m(0) + (1 - \frac{m}{n})\bar{Y}_u(0) = \\ &= \frac{1}{n} \sum_{i=1}^n Y_i(0) = \bar{Y}_n(0) , \end{aligned}$$

então

$$\hat{\theta}^* = \bar{Y}_n(0) \quad (4.4)$$

e

$$\text{Var}_\zeta[\hat{\theta}^*] = \frac{\sigma_0^2}{n} \quad (4.5)$$

o que significa que as observações da segunda ocasião em nada contribuem para o estimador de $\theta(0)$.

b) Objetivo: obter o estimador ζ_c -ótimo para $E_\zeta[\theta(1)]$. Neste caso:

$$c(t) = \begin{cases} 1/p_1 & p|t = 1 \\ 0 & cc \end{cases}$$

Usando o mesmo raciocínio do item anterior:

$$\hat{\theta}^* = \rho K(1-\lambda) (\bar{Y}_u(0) - \bar{Y}_m(0)) + \bar{Y}_m(1) \quad (4.6)$$

e

$$\text{Var}_{\zeta} [\hat{\theta}^*] = \frac{\sigma^2}{m} [1 - \rho^2(1-\lambda)] \quad (4.7)$$

A derivação desses resultados encontram-se no resultado 10 do apêndice.

Se $\rho = 0$, as observações da 1ª ocasião não contribuem para o estimador e $\hat{\theta}^*$ se reduz a $\bar{Y}_m(1)$.

Se $\rho \neq 0$, então $\text{Var}_{\zeta} [\hat{\theta}^*] < \text{Var}_{\zeta} [\bar{Y}_m(1)]$, isto é, o estimador utiliza as observações do instante "0" como informação adicional, para "melhorar" o estimador $\bar{Y}_m(1)$.

Por outro lado, pode-se obter $\hat{\theta}^*$ da seguinte maneira: estimar cada uma das observações perdidas através da regressão, usando o conhecimento fornecido pelos pares completos e depois, calcular a média de todos os valores no instante 1 (m valores conhecidos e u valores estimados).

Assim,

$$\hat{Y}_i(1) = \bar{Y}_m(1) + \beta(Y_i(0) - \bar{Y}_m(0)) \quad \text{onde} \quad \beta = \rho K$$

$$\begin{aligned} \hat{\theta}^* &= \frac{1}{n} \left[\sum_{i=1}^m Y_i(1) + \sum_{i=m+1}^n \hat{Y}_i(1) \right] = \\ &= \frac{1}{n} \left[m\bar{Y}_m(1) + \sum_{i=m+1}^n \left[\bar{Y}_m(1) + \rho K Y_i(0) - \rho K \bar{Y}_m(0) \right] \right] = \\ &= \lambda \bar{Y}_m(1) + \frac{1}{n} \left[u\bar{Y}_m(1) + \rho K u \bar{Y} (0) - \rho K u \bar{Y}_m(0) \right] = \end{aligned}$$

$$\begin{aligned}
 &= \lambda \bar{Y}_m(1) + \frac{u}{n} \left(\bar{Y}_m(1) + \rho_K (\bar{Y}_u(0) - \bar{Y}_m(0)) \right) = \\
 &= \lambda \bar{Y}_m(1) + (1-\lambda) \bar{Y}_m(1) + (1-\lambda) \rho_K (\bar{Y}_u(0) - \bar{Y}_m(0)) = \\
 &= \bar{Y}_m(1) + \rho_K (1-\lambda) (\bar{Y}_u(0) - \bar{Y}_m(0)) = \hat{\theta}^*
 \end{aligned}$$

Como $\hat{\theta}' = \hat{\theta}^*$, pode-se entender intuitivamente, através de $\hat{\theta}'$, o significado de $\hat{\theta}^*$.

c) Objetivo: Obter o estimador ζ_c -ótimo para

$$E_{\zeta}[\theta(1)] - E_{\zeta}[\theta(0)].$$

Neste caso,

$$c(t) = \begin{cases} -1/p_0 & p|t = 0 \\ 1/p_1 & p|t = 1 \\ 0 & cc \end{cases}$$

Pelo Teorema 4.1, $\hat{\theta}^* = \int_0^T c(t) \hat{\theta}^*(\mathcal{D}) dF(t)$. Neste caso,

$$\hat{\theta}^* = \sum_{i=0}^1 c(i) \hat{\theta}_i^*(\mathcal{D}) p_i = \hat{\theta}_1^*(\mathcal{D}) - \hat{\theta}_0^*(\mathcal{D})$$

Como $\hat{\theta}_0^*(\mathcal{D})$ e $\hat{\theta}_1^*(\mathcal{D})$ já foram encontrados nos itens anteriores, páginas 55 e 56, basta calcular a diferença. Então:

$$\begin{aligned}
 \hat{\theta}^* &= \rho K(1-\lambda) (\bar{Y}_u(0) - \bar{Y}_m(0)) + \bar{Y}_m(1) - \bar{Y}_n(0) = \\
 &= \rho K(1-\lambda) \bar{Y}_u(0) - \rho K(1-\lambda) \bar{Y}_m(0) + \bar{Y}_m(1) - (1-\lambda) \bar{Y}_u(0) - \lambda \bar{Y}_m(0) = \\
 &= \bar{Y}_m(1) + [\rho K(1-\lambda) - (1-\lambda)] \bar{Y}_u(0) - [\rho K(1-\lambda) + \lambda] \bar{Y}_m(0) = \\
 &= \bar{Y}_m(1) - [1 - (\lambda + \rho K(1-\lambda))] \bar{Y}_u(0) - [\lambda + \rho K(1-\lambda)] \bar{Y}_m(0).
 \end{aligned}$$

ou

$$\hat{\theta}^* = \bar{Y}_m(1) - (1-\beta) \bar{Y}_u(0) - \beta \bar{Y}_m(0), \quad (4.8)$$

onde

$$\beta = \lambda + \rho K(1-\lambda) \quad (4.9)$$

Neste caso, pode-se também calcular a variância de $\hat{\theta}^*$. Para isso, $\bar{Y}_m(1) - \bar{Y}_m(0)$ será considerado como o estimador não viciado de $E_{\zeta}[\theta(1) - \theta(0)]$. Aplicando-se o teorema II do apêndice:

$$\begin{aligned}
 \text{Var}_{\zeta}[\hat{\theta}^*] &= \text{cov}_{\zeta}[\hat{\theta}^*, \bar{Y}_m(1) - \bar{Y}_m(0)] = \\
 &= \text{cov}_{\zeta}[\bar{Y}_m(1) - (1-\beta) \bar{Y}_u(0) - \beta \bar{Y}_m(0), \bar{Y}_m(1) - \bar{Y}_m(0)] = \\
 &= \frac{1}{m} \sigma_1^2 - \frac{1}{m} \rho \sigma_0 \sigma_1 - \beta \frac{1}{m} \rho \sigma_0 \sigma_1 + \beta \frac{1}{m} \sigma_0^2 = \\
 &= \frac{\sigma_0^2}{m} (K^2 - \rho K - \beta \rho K + \beta),
 \end{aligned}$$

ou substituindo o valor de β ,

$$\text{Var}_{\zeta} [\hat{\theta}^*] = \frac{\sigma_0^2}{m} (K^2 + \lambda (1-2\rho K) - (1-\lambda)\rho^2 K^2). \quad (4.10)$$

II - Painel rotativo

O conjunto de dados será representado por

$$\begin{array}{ccccccc} y_1(0) & \dots & y_m(0) & y_{m+1}(0) & \dots & y_n(0) & \\ y_1(1) & \dots & y_m(1) & & & y_{n+1}(1) & \dots y_{n+u}(1) \end{array}$$

Neste caso, m elementos responderam à pesquisa nas duas ocasiões "0" e "1" e u = n-m elementos responderam somente para uma das duas ocasiões.

- a) Objetivo: estimar a esperança, com relação a ζ , da média populacional no instante "0", $E_{\zeta}[\theta(0)]$.

A matriz de dados, $\tilde{P}$, será dada por

$$\tilde{P} = \begin{pmatrix} y_1(0) \dots y_m(0) & y_{m+1}(0) & \dots & y_n(0) \\ y_1(1) \dots y_m(1) & & & y_{n+1}(1) \dots y_{n+u}(1) \end{pmatrix}$$

Pelo Teorema I do apêndice, $\hat{\theta}^*$ terá a seguinte forma:

$$\begin{aligned}\hat{\theta}^* &= a'y_1(0) + \dots + a'y_m(0) + b'y_{m+1}(0) + \dots + b'y_n(0) + \\ &+ c'y_1(1) + \dots + c'y_m(1) + d'y_{n+1}(1) + \dots + d'y_{n+u}(1) = \\ &= a' \sum_{i=1}^m y_i(0) + b' \sum_{i=m+1}^n y_i(0) + c' \sum_{i=1}^m y_i(1) + d' \sum_{i=n+1}^{n+u} y_i(1)\end{aligned}$$

Usando $\bar{y}_m(0)$, $\bar{y}_u(0)$, $\bar{y}_m(1)$, já especificados anteriormente, e

$$\bar{y}_u(1) = \frac{1}{u} \sum_{i=n+1}^{n+u} y_i(1),$$

$$\hat{\theta}^* = a\bar{y}_m(0) + b\bar{y}_u(0) + c\bar{y}_m(1) + d\bar{y}_u(1).$$

Para que $\hat{\theta}^*$ seja não viciado para $E_{\zeta}[\theta]$, é preciso escolher a , b , c e d tais que $a+b = 1$ e $c+d = 0$, ou seja,

$$\hat{\theta}^* = a\bar{y}_m(0) + (1-a)\bar{y}_u(0) + c\bar{y}_m(1) - c\bar{y}_u(1).$$

Usando o Teorema II (apêndice), $\text{Cov}_{\zeta}[\hat{\theta}^*, \hat{\theta}_0] = 0$,

$$\begin{aligned}\text{cov}_{\zeta}[\hat{\theta}^*, \hat{\theta}_0] &= \text{Cov}_{\zeta}[a\bar{y}_m(0) + (1-a)\bar{y}_u(0) + c\bar{y}_m(1) - c\bar{y}_u(1), \\ &\quad \bar{y}_m(0) - \bar{y}_u(0)] = \\ &= a \frac{1}{m} \sigma_0^2 - (1-a) \frac{1}{u} \sigma_0^2 + c \frac{1}{m} \rho \sigma_0 \sigma_1 - c \frac{1}{u} \rho \sigma_0 \sigma_1 = 0\end{aligned}$$

$$a_u \sigma_0^2 - m \sigma_0^2 + a m \sigma_0^2 + c \rho \sigma_0 \sigma_1 u = 0$$

$$a \sigma_0^2 (u+m) + c \rho \sigma_0 \sigma_1 u = m \sigma_0^2 \quad (4.11)$$

Por outro lado, pode-se considerar como estimador de zero, $\bar{y}_m(1) - \bar{y}_u(1)$, e, aplicando-se novamente o Teorema II

(apêndice), tem-se:

$$\begin{aligned}
 \text{Cov}_{\zeta} [a\bar{Y}_m(0) + (1-a)\bar{Y}_u(0) + c\bar{Y}_m(1) - c\bar{Y}_u(1), \bar{Y}_m(1) - \bar{Y}_u(1)] &= \\
 = a \frac{1}{m} \rho \sigma_0 \sigma_1 + c \frac{1}{m} \sigma_1^2 + c \frac{1}{u} \sigma_1^2 &= 0 \\
 a \rho \sigma_0 \sigma_1 + c \sigma_1^2 u + c \sigma_1^2 m &= 0 \\
 a \rho \sigma_0 \sigma_1 u + c \sigma_1^2 n &= 0 \tag{4.12}
 \end{aligned}$$

Resolvendo o sistema composto pelas equações (4.11) e (4.12), tem-se

$$\begin{aligned}
 a &= \frac{-c \sigma_1^2 n}{\rho \sigma_0 \sigma_1 u} = \frac{-c K}{\rho (1-\lambda)} \\
 \frac{-c K}{\rho (1-\lambda)} \sigma_0^2 (u+m) + c \rho \sigma_0 \sigma_1 u &= m \sigma_0^2 \\
 c (-K \sigma_0 n + \rho^2 \sigma_1 u (1-\lambda)) &= m \sigma_0 \rho (1-\lambda) \\
 c &= \frac{m \sigma_0 \rho (1-\lambda)}{\sigma_1 (u (1-\lambda) \rho^2 - n)} = \frac{-\lambda (1-\lambda) \rho / K}{1 - (1-\lambda)^2 \rho^2} \\
 a &= \frac{-c K}{\rho (1-\lambda)} = \frac{\lambda (1-\lambda) \rho}{\rho (1-\lambda) [1 - (1-\lambda)^2 \rho^2]} = \frac{\lambda}{1 - (1-\lambda)^2 \rho^2} \\
 \text{ou seja,} \\
 \hat{\theta}^* &= \frac{1}{1 - (1-\lambda)^2 \rho^2} (\lambda \bar{Y}_m(0) - \lambda \bar{Y}_u(0) - \lambda (1-\lambda) \frac{\rho}{K} \bar{Y}_m(1) + \lambda (1-\lambda) \frac{\rho}{K} \bar{Y}_u(1)) + \bar{Y}_u(0) \tag{4.13}
 \end{aligned}$$

Considerando $\bar{Y}_u(0)$, que é outro estimador não viciado de $E_\zeta[\theta]$ e aplicando o Teorema II (apêndice):

$$\begin{aligned} \text{Var}_\zeta[\hat{\theta}^*] &= \text{cov}_\zeta[\hat{\theta}^*, \bar{Y}_u(0)] = \\ &= \text{cov}_\zeta[a\bar{Y}_m(0) + (1-a)\bar{Y}_u(0) + c\bar{Y}_m(1) - c\bar{Y}_u(1), \bar{Y}_u(0)] = \\ &= (1-a)\frac{1}{u}\sigma_0^2 = \frac{1-(1-\lambda)^2\rho^2 - \lambda}{1-(1-\lambda)^2\rho^2} \frac{\sigma_0^2}{u} = \frac{\sigma_0^2(1-\lambda)[1-(1-\lambda)\rho^2]}{[1-(1-\lambda)^2\rho^2]u} \end{aligned}$$

ou

$$\text{Var}_\zeta[\hat{\theta}^*] = \frac{\sigma_0^2}{n} \left(\frac{1-(1-\lambda)\rho^2}{1-(1-\lambda)^2\rho^2} \right) \quad (4.14)$$

Considerações

1) Se $\lambda = 0$, todos os elementos da amostra na primeira ocasião foram substituídos por novos elementos, ou seja, $u = n$. Então,

$$\hat{\theta}^* = \bar{Y}_n(0) \quad \text{e} \quad \text{Var}_\zeta[\hat{\theta}^*] = \frac{\sigma_0^2}{n}.$$

2) Se $\lambda = 1$, todos os elementos da amostra na primeira ocasião foram mantidos para a segunda ocasião, ou seja, $m = n$ (painel discreto completo). Novamente,

$$\hat{\theta}^* = \bar{Y}_n(0) \quad \text{e} \quad \text{Var}_\zeta[\hat{\theta}^*] = \frac{\sigma_0^2}{n}.$$

3) Se $0 < \lambda < 1$,

$$\hat{\theta}^* = A\bar{Y}_m(0) + (1-A)\bar{Y}_u(0) + B\bar{Y}_m(1) - B\bar{Y}_u(1) \quad (4.15)$$

$$e \quad \text{Var}_{\zeta} [\hat{\theta}^*] = \frac{\sigma_0^2}{n} R, \quad (4.16)$$

$$\text{onde} \quad A = \frac{\lambda}{1 - (1-\lambda)^2 \rho^2}, \quad B = \frac{-\lambda(1-\lambda)\rho/K}{1 - (1-\lambda)^2 \rho^2}$$

$$e \quad R = \frac{1 - (1-\lambda)\rho^2}{1 - (1-\lambda)^2 \rho^2} \quad (4.17)$$

$$\text{Neste caso, } \text{Var}_{\zeta} [\hat{\theta}^*] < \frac{\sigma_0^2}{n} \text{ pois } R < 1.$$

b) Objetivo: estimar a $E_{\zeta} [\theta(1)]$.

Pelo mesmo raciocínio anterior, conforme derivação no apêndice (resultado 11),

$$\begin{aligned} \hat{\theta}^* = \frac{1}{1 - (1-\lambda)^2 \rho^2} & (-\lambda \rho K(1-\lambda) \bar{Y}_m(0) + \lambda \rho K(1-\lambda) \bar{Y}_u(0) + \lambda \bar{Y}_m(1) - \lambda \bar{Y}_u(1)) + \\ & + \bar{Y}_u(1) \end{aligned} \quad (4.18)$$

$$e \quad \text{Var}_{\zeta} [\hat{\theta}^*] = \frac{\sigma_1^2}{n} \left(\frac{1 - (1-\lambda)\rho^2}{1 - (1-\lambda)^2 \rho^2} \right) \quad (4.19)$$

Considerações:

1) Se $\lambda = 0$,

$$\hat{\theta}^* = \bar{Y}_n(1) \quad e \quad \text{Var}_{\zeta} [\hat{\theta}^*] = \frac{\sigma_1^2}{n}$$

2) Se $\lambda = 1$,

$$\hat{\theta}^* = \bar{Y}_n(1) \quad e \quad \text{Var}_{\zeta} [\hat{\theta}^*] = \frac{\sigma_1^2}{n}$$

3) Se $0 < \lambda < 1$

$$\hat{\theta}^* = C\bar{Y}_m(0) - C\bar{Y}_u(0) + A\bar{Y}_m(1) + (1-A)\bar{Y}_u(1) \quad (4.20)$$

e

$$\text{Var}_{\zeta} [\hat{\theta}^*] = \frac{\sigma_1^2}{n} R \quad (4.21)$$

onde A e R foram dados no item anterior e $C = \frac{-\lambda \rho_K (1-\lambda)}{1 - (1-\lambda)^2 \rho^2}$.

Também aqui $\text{Var}_{\zeta} [\hat{\theta}^*] < \frac{\sigma_1^2}{n}$ pois $R < 1$.

c) Objetivo: estimar a variação entre as esperanças das médias populacionais para os dois instantes, $E_{\zeta} [\theta(1)] - E_{\zeta} [\theta(0)]$.

Pelo Teorema 4.1, analogamente ao caso anterior,

$$\hat{\theta}^* = \hat{\theta}_1^*(\mathcal{D}) - \hat{\theta}_0^*(\mathcal{D})$$

$$\begin{aligned} \hat{\theta}^* &= C\bar{Y}_m(0) - C\bar{Y}_u(0) + A\bar{Y}_m(1) + (1-A)\bar{Y}_u(1) - \\ &\quad - [A\bar{Y}_m(0) + (1-A)\bar{Y}_u(0) + B\bar{Y}_m(1) - B\bar{Y}_u(1)] = \\ &= (C-A)\bar{Y}_m(0) - (1+C-A)\bar{Y}_u(0) + (A-B)\bar{Y}_m(1) + (1-A+B)\bar{Y}_u(1). \end{aligned}$$

Para calcular a variância de $\hat{\theta}^*$, será considerado $\bar{Y}_u(1) - \bar{Y}_u(0)$ como o outro estimador não viciado de $E_{\zeta} [\theta]$. Então

$$\begin{aligned} \text{Var}_{\zeta} [\hat{\theta}^*] &= \text{cov}_{\zeta} [(C-A)\bar{Y}_m(0) - (1+C-A)\bar{Y}_u(0) + (A-B)\bar{Y}_m(1) + \\ &\quad + (1-A+B)\bar{Y}_u(1), \bar{Y}_u(1) - \bar{Y}_u(0)] = \\ &= (1+C-A) \frac{1}{u} \sigma_0^2 + (1-A+B) \frac{1}{u} \sigma_1^2, \end{aligned}$$

substituindo os valores de A, B e C:

$$\begin{aligned}
 \text{Var}_{\zeta}[\hat{\theta}^*] &= \frac{1 - (1-\lambda)^2 \rho^2 - \lambda \rho K (1-\lambda) - \lambda}{u(1 - (1-\lambda)^2 \rho^2)} \sigma_0^2 + \\
 &+ \frac{1 - (1-\lambda)^2 \rho^2 - \lambda - \lambda(1-\lambda) \rho/K}{u(1 - (1-\lambda)^2 \rho^2)} \sigma_1^2 = \\
 &= \frac{(1-\lambda)(1 - (1-\lambda) \rho^2) \sigma_0^2 - \lambda(1-\lambda) \rho \sigma_0 \sigma_1 + (1-\lambda)(1 - (1-\lambda) \rho^2) \sigma_1^2 - \lambda(1-\lambda) \rho \sigma_0 \sigma_1}{n(1-\lambda)(1 - (1-\lambda)^2 \rho^2)} = \\
 &= \frac{(1 - (1-\lambda) \rho^2) \sigma_0^2 + (1 - (1-\lambda) \rho^2) \sigma_1^2 - 2\lambda \rho \sigma_0 \sigma_1}{n(1 - (1-\lambda)^2 \rho^2)} = \\
 &= \frac{1 - (1-\lambda) \rho^2}{1 - (1-\lambda)^2 \rho^2} \left(\frac{\sigma_0^2 + \sigma_1^2}{n} \right) - \frac{2\lambda}{1 - (1-\lambda)^2 \rho^2} \frac{\rho \sigma_0 \sigma_1}{n} \quad (4.22)
 \end{aligned}$$

Considerações:

1) Se $\lambda = 0$, $u = n$,

$$\hat{\theta}^* = \bar{Y}_n(1) - \bar{Y}_n(0) \quad \text{e} \quad \text{Var}_{\zeta}[\hat{\theta}^*] = \frac{\sigma_0^2}{n} + \frac{\sigma_1^2}{n}$$

2) Se $\lambda = 1$, $m = n$,

$$\hat{\theta}^* = \bar{Y}_n(1) - \bar{Y}_n(0) \quad \text{e} \quad \text{Var}_{\zeta}[\hat{\theta}^*] = \frac{\sigma_0^2}{n} + \frac{\sigma_1^2}{n} - \frac{2\rho \sigma_0 \sigma_1}{n}$$

3) Se $0 < \lambda < 1$,

$$\hat{\theta}^* = (C-A) \bar{Y}_m(0) - (1+C-A) \bar{Y}_u(0) + (A-B) \bar{Y}_m(1) + (1-A+B) \bar{Y}_u(1) \quad (4.23)$$

e

$$\text{Var}_{\zeta}[\hat{\theta}^*] = R \left(\frac{\sigma_0^2}{n} + \frac{\sigma_1^2}{n} \right) - \frac{2\lambda}{1 - (1-\lambda)^2 \rho^2} \frac{\rho \sigma_0 \sigma_1}{n}, \quad (4.24)$$

onde A, B, C e R são os mesmos dos itens anteriores.

4.4. Aplicação II - Caso Discreto (várias ocasiões)

Nesta seção, será suposto que os n elementos da amostra respondem à pesquisa no instante "0" e, m desses n elementos respondem também para T ocasiões distintas. Os dados poderão ser representados por:

$$\begin{array}{ccccccc} y_1(0) & \dots\dots & y_m(0) & y_{m+1}(0) & \dots\dots & y_n(0) \\ y_1(1) & \dots\dots & y_m(1) & & & \\ y_1(2) & \dots\dots & y_m(2) & & & \\ \vdots & & & & & \\ y_1(T) & \dots\dots & y_m(T) & & & \end{array}$$

Considerando $F(t)$ como a distribuição uniforme discreta nas $T+1$ ocasiões, haverá interesse em estimar $E_{\zeta}[\theta]$, para as seguintes situações:

- a) $c(t) = 1$, para $\forall t \in P_2(T)$
- b) $c(t) = T+1$, para $\forall t \in P_2(T)$
- c) $c(t) = N$, para $\forall t \in P_2(T)$
- d) $c(t) = N(T+1)$, para $\forall t \in P_2(T)$

$$dF(t) = \frac{1}{T+1} , \forall t \in P_2(T)$$

A fim de estimar $E_{\zeta}[\theta]$, é necessário primeiro estimar $E_{\zeta}[\theta(t)]$, para $t \in P_2(T)$. Considerando a matriz de covariância

$$\Sigma = \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 & \rho\sigma_0\sigma_2 & \dots & \rho\sigma_0\sigma_T \\ \rho\sigma_0\sigma_1 & \sigma_1^2 & \rho\sigma_1\sigma_2 & \dots & \rho\sigma_1\sigma_T \\ \vdots & \vdots & \vdots & & \vdots \\ \rho\sigma_0\sigma_T & \rho\sigma_1\sigma_T & \rho\sigma_2\sigma_T & \dots & \sigma_T^2 \end{pmatrix} ,$$

o estimador ótimo de $E_{\zeta}[\theta(t)]$, obtido através do mesmo critério da seção anterior, é dado por:

$$\hat{\theta}_t^* = \rho K(t) (1-\lambda) (\bar{Y}_u(0) - \bar{Y}_m(0)) + \bar{Y}_m(t) ,$$

onde $K(t) = \frac{\sigma_t}{\sigma_0}$.

Aplicando o Teorema 4.1, pode-se obter o estimador $\hat{\theta}$ ótimo para $E_{\zeta}[\theta]$. Os resultados, para as quatro situações de interesse, são os seguintes:

a) $c(t) = 1 \quad E_{\zeta}[\theta] = \frac{1}{T+1} \sum_{i=0}^T \frac{1}{N} E_{\zeta}[1'Y(i)]$

θ é a média das $(T+1)$ médias

$$\hat{\theta}^* = \rho(1-\lambda) (\bar{Y}_u(0) - \bar{Y}_m(0)) \sum_{i=0}^T k(i) \frac{1}{T+1} + \sum_{i=0}^T \bar{Y}_m(i) \frac{1}{T+1}$$

se $K(i) = K$ (constante),

$$\hat{\theta}^* = \rho(1-\lambda) (\bar{Y}_u(0) - \bar{Y}_m(0)) K + \frac{1}{T+1} \sum_{i=0}^T \bar{Y}_m(i)$$

além disso, se $\lambda = 1$ (painel completo),

$$\hat{\theta}^* = \frac{1}{T+1} \sum_{i=0}^T \bar{Y}_n(i)$$

$$b) \quad c(t) = T+1 \quad E_{\zeta}[\theta] = \sum_{i=0}^T \frac{1}{N} E_{\zeta}[\mathbf{1}'_{\sim N} \mathbf{Y}(i)]$$

θ é a soma das $(T+1)$ médias

$$\hat{\theta}^* = \rho(1-\lambda) (\bar{Y}_u(0) - \bar{Y}_m(0)) \sum_{i=0}^T K(i) + \sum_{i=0}^T \bar{Y}_m(i)$$

se $K(i) = K$,

$$\hat{\theta}^* = \rho(1-\lambda) (\bar{Y}_u(0) - \bar{Y}_m(0)) K(T+1) + \sum_{i=0}^T \bar{Y}_m(i)$$

se $\lambda = 1$,

$$\hat{\theta}^* = \sum_{i=0}^T \bar{Y}_n(i).$$

$$c) \quad c(t) = N \quad E_{\zeta}[\theta] = \frac{1}{T+1} \sum_{i=0}^T E_{\zeta}[\mathbf{1}'_{\sim N} \mathbf{Y}(i)]$$

θ é a média dos $(T+1)$ totais

$$\hat{\theta}^* = \rho(1-\lambda) (\bar{Y}_u(0) - \bar{Y}_m(0)) \sum_{i=0}^T N K(i) \frac{1}{T+1} + \sum_{i=0}^T N \bar{Y}_m(i) \frac{1}{T+1}$$

se $K(i) = K$,

$$\hat{\theta}^* = \rho(1-\lambda) (\bar{Y}_u(0) - \bar{Y}_m(0)) N \cdot K + \frac{N}{T+1} \sum_{i=0}^T \bar{Y}_m(i)$$

$$\text{se } \lambda = 1, \quad \hat{\theta}^* = \frac{N}{T+1} \sum_{i=0}^T \bar{Y}_n(i)$$

$$d) \quad c(t) = N(T+1) \quad E_{\zeta}[\theta] = \sum_{i=0}^T E_{\zeta} \left[\frac{1}{N} Y(i) \right]$$

θ é a soma dos $(T+1)$ totais

$$\hat{\theta}^* = \rho(1-\lambda) (\bar{Y}_u(0) - \bar{Y}_m(0)) N \sum_{i=0}^T K(i) + N \sum_{i=0}^T \bar{Y}_m(i)$$

se $K(i) = K$,

$$\hat{\theta}^* = \rho(1-\lambda) (\bar{Y}_u(0) - \bar{Y}_m(0)) N(T+1)K + N \sum_{i=0}^T \bar{Y}_m(i)$$

$$\text{se } \lambda = 1, \quad \hat{\theta}^* = N \sum_{i=0}^T \bar{Y}_n(i)$$

4.5. Aplicação III - Caso Contínuo

Nesta seção, o conjunto de dados será representado por:

ou seja n elementos respondem à pesquisa no instante "0", mas somente m desses n tem a sua trajetória acompanhada durante o intervalo $[0, T]$.

Se $\lambda = 1$, a trajetória dos n elementos é conhecida para todo o intervalo de tempo.

Situações análogas ao caso discreto serão consideradas, ou seja, $F(t)$ é a distribuição uniforme no intervalo $[0, T]$ e $c(t)$ é, para $\forall t \in P_2(T)$, o seguinte:

$$a) \quad c(t) = 1, \quad E_{\zeta}[\theta] = \frac{1}{T} \int_0^T \frac{1}{N} E_{\zeta}[\mathbf{1}'_{\mathbf{N}} \mathbf{Y}(t)] dt$$

$E_{\zeta}[\theta]$ seria a média das médias, com relação ao intervalo $[0, T]$.

Logo,

$$\hat{\theta}^* = \rho(1-\lambda) (\bar{Y}_u(0) - \bar{Y}_m(0)) \frac{1}{T} \int_0^T K(t) dt + \frac{1}{T} \int_0^T \bar{Y}_m(t) dt$$

se $K(t) = K$,

$$\hat{\theta}^* = \rho(1-\lambda) (\bar{Y}_u(0) - \bar{Y}_m(0)) K + \frac{1}{T} \int_0^T \bar{Y}_m(t) dt$$

$$\text{se } \lambda = 1, \quad \hat{\theta}^* = \frac{1}{T} \int_0^T \bar{Y}_n(t) dt$$

$$b) \quad c(t) = T, \quad E_{\zeta}[\theta] = \int_0^T \frac{1}{N} E_{\zeta}[\mathbf{1}'_{\mathbf{N}} \mathbf{Y}(t)] dt$$

o significado de $E_{\zeta}[\theta]$ seria o de acompanhar o comportamento da média ao longo do tempo. Logo,

$$\hat{\theta}^* = \rho(1-\lambda)(\bar{Y}_u(0) - \bar{Y}_m(0)) \int_0^T K(t) dt + \int_0^T \bar{Y}_m(t) dt$$

se $K(t) = K$,

$$\hat{\theta}^* = \rho(1-\lambda)(\bar{Y}_u(0) - \bar{Y}_m(0)) K.T + \int_0^T \bar{Y}_m(t) dt$$

se $\lambda = 1$,

$$\hat{\theta}^* = \int_0^T \bar{Y}_n(t) dt$$

c) $c(t) = N$ $E_{\zeta}[\theta] = \frac{1}{T} \int_0^T E_{\zeta}[1_N' Y(t)] dt$

$E_{\zeta}[\theta]$ seria a média, em relação ao tempo, das médias dos totais. Logo,

$$\hat{\theta}^* = \rho(1-\lambda)(\bar{Y}_u(0) - \bar{Y}_m(0)) \frac{N}{T} \int_0^T K(t) dt + \frac{N}{T} \int_0^T \bar{Y}_m(t) dt$$

se $K(t) = K$,

$$\hat{\theta}^* = \rho(1-\lambda)(\bar{Y}_u(0) - \bar{Y}_m(0)) N.K + \frac{N}{T} \int_0^T \bar{Y}_m(t) dt$$

se $\lambda = 1$,

$$\hat{\theta}^* = \frac{N}{T} \int_0^T \bar{Y}_n(t) dt$$

d) $c(t) = TN$ $E_{\zeta}[\theta] = \int_0^T N 1_N' E_{\zeta}[Y(t)] dt$

o significado de $E_{\zeta}[\theta]$ seria o de acompanhar o comportamento da média do total, em relação a ζ , ao longo do tempo. Portanto,

$$\hat{\theta}^* = \rho(1-\lambda)(\bar{Y}_u(0) - \bar{Y}_m(0))N \int_0^T K(t) dt + N \int_0^T \bar{Y}_m(t) dt$$

se $K(t) = K$, ou

$$\hat{\theta}^* = \rho(1-\lambda)(\bar{Y}_u(0) - \bar{Y}_m(0))NKT + N \int_0^T \bar{Y}_m(t) dt$$

Note que se $\lambda=1$, então

$$\hat{\theta}^* = N \int_0^T \bar{Y}_n(t) dt.$$

oOo

CAPÍTULO 5

OBTENÇÃO DE PREVISORES BAYESIANOS ATRAVÉS DO MODELO DE FILTRO DE KALMAN

5.1. Introdução

Neste capítulo, o modelo de Filtro de Kalman será utilizado para a obtenção de previsores Bayesianos para θ . Será considerado somente o caso em que $F(t)$ é uma função distribuição associada a uma variável aleatória discreta, ou seja, θ se reduz a

$$\theta = \sum_{j=0}^T c(j)\theta(j)p_j .$$

Por outro lado, serão consideradas duas alternativas para $\theta(t)$: a média populacional no instante t , $\theta(t) = \frac{1}{N} \mathbf{1}' \mathbf{Y}(t)$, e o total populacional no instante t , $\theta(t) = \mathbf{1}' \mathbf{Y}(t)$.

O objetivo será o de prever θ , de maneira recursiva, adotando o modelo de Superpopulação Normal e supondo que uma combinação linear das médias (totais) populacionais segue o modelo de Filtro de Kalman (para mais informações sobre o Filtro de Kalman ver Morettin/Toloi (1981)).

Com relação às variâncias, serão mencionados dois casos: variâncias conhecidas e desconhecidas.

Será denotada por θ_t , a combinação dos $\theta(j)$'s que leva em conta os instantes de tempo $1, 2, \dots, t$, ou seja,

$$\theta_t = \sum_{j=1}^t c(j)\theta(j)p_j.$$

Note que, neste capítulo, para simplificação das fórmulas, o primeiro instante foi denotado por 1 e não por zero.

Como já foi dito anteriormente, os $c(j)$'s e p_j 's serão especificados de acordo com o interesse do pesquisador. Como $F(t)$ corresponde à variável discreta, os p_j 's serão valores positivos fixados, associados ao instante j e, dessa forma, pode-se escrever $c(j)p_j = \ell_j$ e $\theta_t = \sum_{j=1}^t \ell_j \theta(j)$. Por consequência, os ℓ_j 's também estarão especificados.

Para o caso de variâncias conhecidas, a abordagem Bayesiana adotada neste capítulo, será um caso particular daquela adotada por Meinhold/Singpurwalla (1983) e, para o caso de variâncias desconhecidas, de Broemeling/Diaz/Yusoff (1985). (Ver Bolfarine/Rodrigues (87)).

5.2. Modelo de Filtro de Kalman com variâncias conhecidas

Suponha que $\underline{Y}(t)$ tenha distribuição normal multivariada com média $\mu \underline{1}_N'$ e matriz de covariâncias $\sigma^2 \underline{I}$.

Além disso, lembrando que $P_1(t) = \{1, 2, \dots, N\}$, suponha que observações correspondentes a elementos diferentes k e $k' \in P_1(t)$ sejam não correlacionadas para qualquer instante, e que observações correspondentes a um mesmo elemento, para instantes diferentes tenham correlação ρ , isto é,

$$\underline{Y}(t) \sim N(\mu \underline{1}_N', \sigma^2 \underline{I})$$

e

$$\text{Cov}[Y_k(t), Y_{k'}(t)] = \begin{cases} \sigma^2 \rho & , \quad k=k' \quad , \quad t \neq t' \\ 0 & , \quad k \neq k' \quad , \quad t \neq t' \end{cases}$$

onde σ^2 e ρ são conhecidos e μ é desconhecido.

Uma amostra de n elementos será retirada da população para cada instante, de modo que o conjunto de dados seja um painel rotativo.

Caso 1: $\theta(t)$ é a média populacional no instante t ,

$$\theta(t) = \frac{1}{N} \sum_{i=1}^N Y_i(t) .$$

Seja $\hat{\theta}_t = \sum_{j=1}^t \lambda_j \hat{\theta}(j)$, onde $\hat{\theta}(j)$ é a média amostral

no instante j , ou seja,

$$\hat{\theta}(j) = \frac{1}{n} \sum_{k \in S_j} Y_k(j)$$

onde S_j é a amostra no instante j .

Utilizando o resultado 5 do apêndice, tem-se que:

$$E[\hat{\theta}_t / \theta_t] = \theta_t$$

e

$$\text{Var}[\hat{\theta}_t / \theta_t] = A_t \sigma^2$$

onde

$$A_t = \frac{1}{n} (1-f) \sum_{j=1}^t \lambda_j^2 + \frac{2\rho}{n} \sum_{j < i} \lambda_j \lambda_i (\lambda_{ij} - f) ,$$

$f = \frac{n}{N}$ (fração amostral), $\lambda_{ij} = \frac{m_{ij}}{n}$; m_{ij} é o número de elementos que estão na amostra nos tempos i e j simultaneamente, ou seja, λ_{ij} é a proporção de elementos comuns aos instantes i e j , em relação ao tamanho da amostra (fixo) em cada instante.

Caso 2: $\theta(t)$ é o total populacional no instante t ,

$$\theta(t) = \frac{1}{N} \sum_{j=1}^t Y_j(t)$$

Novamente $\hat{\theta}_t = \sum_{j=1}^t \lambda_j \theta(t)$, só que neste caso $\hat{\theta}(t)$ é o estimador expansão, ou seja,

$$\hat{\theta}(t) = \frac{N}{n} \sum_{k \in S_j} Y_k(j).$$

Usando o mesmo raciocínio do caso 1, não é difícil verificar que:

$$E[\hat{\theta}_t | \theta_t] = \theta_t$$

e

$$\text{Var}[\hat{\theta}_t | \theta_t] = N^2 A_t \sigma^2$$

A fim de englobar os dois casos de uma maneira geral, B_t será definido como segue:

$$B_t = \begin{cases} A_t, & \text{se } \theta(t) \text{ é a média populacional} \\ N^2 A_t, & \text{se } \theta(t) \text{ é o total populacional} \end{cases}$$

Assim, pode-se escrever $\hat{\theta}_t$ da seguinte forma:

$$\hat{\theta}_t = \theta_t + e_t,$$

onde os e_t têm distribuição Normal (por suposição) com média zero, variância $B_t \sigma^2$ e são não correlacionados.

Por outro lado, como existe correlação entre elementos da população, para instantes diferentes, pode-se supor que θ_t dependa de θ_{t-1} . Então,

$$\theta_t = \beta \theta_{t-1} + \varepsilon_t$$

onde ε_t tem distribuição Normal (por suposição) com média zero, variância V conhecida e são não correlacionados. Além disso, será suposto que e_t e ε_t são independentes, $\forall t$, e ε_t independente de $Y_k(j)$, $\forall j \leq t-1$, $\forall k \in P_1(t)$.

Reunindo todas essas informações, o modelo de Filtro de Kalman será dado pela equação das observações (1) e pela equação do sistema (2):

$$(1) \quad \hat{\theta}_t = \theta_t + e_t, \quad e_t \sim N(0, B_t \sigma^2)$$

$$(2) \quad \theta_t = \beta \theta_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, V)$$

$B_t \sigma^2$, V e β são conhecidos.

Agora, seja $\mathcal{D}_t = \{(k, Y_k(j)), k \in S_j, j \in \{1, 2, \dots, t\}\}$ o conjunto de dados até o instante t . Usando o teorema de Bayes, pode-se escrever:

$$P(\theta_t | \hat{\theta}_t, \mathcal{D}_{t-1}) = \frac{P(\hat{\theta}_t | \theta_t, \mathcal{D}_{t-1}) P(\theta_t | \mathcal{D}_{t-1})}{P(\hat{\theta}_t | \mathcal{D}_{t-1})},$$

onde,

$P(\theta_t | \hat{\theta}_t, \mathcal{D}_{t-1})$: distribuição a posteriori de θ_t

$P(\hat{\theta}_t | \theta_t, \mathcal{D}_{t-1})$: função de verossimilhança

$P(\theta_t | \mathcal{D}_{t-1})$: distribuição a priori de θ_t .

O objetivo é encontrar o Previsor de Kalman, dado por:

$$\tilde{\theta}_t = E[\theta_t | \hat{\theta}_t, \mathcal{D}_{t-1}].$$

Para isso, é necessário conhecer a distribuição a priori de θ_t e a função de verossimilhança.

Distribuição a priori de θ_t

Suponha que no instante $t-1$, a distribuição a posteriori de θ_{t-1} seja Normal com média $\tilde{\theta}_{t-1}$ e variância Σ_{t-1} , ou seja,

$$\theta_{t-1} | \hat{\theta}_{t-1} \sim N(\tilde{\theta}_{t-1}, \Sigma_{t-1}).$$

Pela equação do sistema (2), tem-se que

$$E[\theta_t | \hat{\theta}_{t-1}] = \beta E[\theta_{t-1} | \hat{\theta}_{t-1}] + E[\epsilon_t | \hat{\theta}_{t-1}] = \beta \tilde{\theta}_{t-1}$$

e

$$\begin{aligned} \text{Var}[\theta_t | \hat{\theta}_{t-1}] &= \beta^2 \text{Var}[\theta_{t-1} | \hat{\theta}_{t-1}] + \text{Var}[\epsilon_t | \hat{\theta}_{t-1}] = \\ &= \beta^2 \Sigma_{t-1} + V. \end{aligned}$$

Logo, a priori de θ_t será dada por:

$$\theta_t | \hat{\theta}_{t-1} \sim N(\beta \tilde{\theta}_{t-1}, \beta^2 \Sigma_{t-1} + V).$$

Função de Verossimilhança

Pela equação do sistema (2), pode-se definir outro estimador para θ_t , a partir de $\tilde{\theta}_{t-1}$, dado por $\hat{\hat{\theta}}_t = \beta \tilde{\theta}_{t-1}$.

Após a observação dos dados no instante t , seja ω_t , o erro de previsão de $\hat{\hat{\theta}}_t$ com relação a $\hat{\theta}_t$, ou seja,

$$\omega_t = \hat{\theta}_t - \hat{\hat{\theta}}_t = \hat{\theta}_t - \beta \tilde{\theta}_{t-1}.$$

Como β e $\tilde{\theta}_{t-1}$ são conhecidos, observar ω_t equivale a observar $\hat{\theta}_t$. Então,

$$P(\theta_t | \hat{\theta}_t, \mathcal{D}_{t-1}) = P(\theta_t | \omega_t, \mathcal{D}_{t-1}) = \frac{P(\omega_t | \theta_t, \mathcal{D}_{t-1}) P(\theta_t | \mathcal{D}_{t-1})}{P(\omega_t | \mathcal{D}_{t-1})}$$

A função de verossimilhança será, agora,

$$P(\omega_t | \theta_t, \mathcal{D}_{t-1}) \text{ e,}$$

$$E[\omega_t | \theta_t, \mathcal{D}_{t-1}] = E[\hat{\theta}_t - \beta \tilde{\theta}_{t-1} | \theta_t, \mathcal{D}_{t-1}] = \theta_t - \beta \tilde{\theta}_{t-1}$$

e

$$\text{Var}[\omega_t | \theta_t, \mathcal{D}_{t-1}] = \text{Var}[\hat{\theta}_t - \beta \tilde{\theta}_{t-1} | \theta_t, \mathcal{D}_{t-1}] = B_t \sigma^2.$$

Usando as propriedades da Distribuição Normal Bivariada (resultado 6 do apêndice) e identificando $X_1 = \omega_t$ e $X_2 = \theta_t$, tem-se que

$$\mu_2 = \beta \tilde{\theta}_{t-1} \text{ e } \Sigma_{22} = \beta^2 \Sigma_{t-1} + V.$$

Identificando os outros elementos,

$$\mu_1 + \frac{\Sigma_{12}(\theta_t - \beta \tilde{\theta}_{t-1})}{\beta^2 \Sigma_{t-1} + V} = \theta_t - \beta \tilde{\theta}_{t-1},$$

de onde resulta que $\mu_1 = 0$ e $\Sigma_{12} = \beta^2 \Sigma_{t-1} + V$.

Analogamente, para a variância,

$$\Sigma_{11} = \frac{(\beta^2 \Sigma_{t-1} + V)^2}{\beta^2 \Sigma_{t-1} + V} = \beta_t \sigma^2 ,$$

de onde tem-se

$$\Sigma_{11} = B_t \sigma^2 + \beta^2 \Sigma_{t-1} + V.$$

Escrevendo na forma de matrizes,

$$\begin{pmatrix} \theta_t \\ \omega_t \end{pmatrix} / \hat{\theta}_{t-1} \sim N \left(\begin{pmatrix} \beta \tilde{\theta}_{t-1} \\ 0 \end{pmatrix} , \begin{pmatrix} R_t & R_t \\ R_t & B_t \sigma^2 + R_t \end{pmatrix} \right) ,$$

onde $R_t = \beta^2 \Sigma_{t-1} + V$.

Mas a distribuição de interesse é $P(\theta_t | \hat{\theta}_t, \mathcal{D}_{t-1})$, que será obtida através da seguinte distribuição bivariada

$$\begin{pmatrix} \theta_t \\ \hat{\theta}_t \end{pmatrix} / \hat{\theta}_{t-1} \sim N \left(\begin{pmatrix} \beta \tilde{\theta}_{t-1} \\ \beta \tilde{\theta}_{t-1} \end{pmatrix} , \begin{pmatrix} R_t & R_t \\ R_t & B_t \sigma^2 + R_t \end{pmatrix} \right) .$$

Então, aplicando novamente as propriedades descritas no apêndice, tem-se:

$$(\theta_t | \hat{\theta}_t, \mathcal{D}_{t-1}) \sim N(\tilde{\theta}_t, \Sigma_t) ,$$

onde

$$\tilde{\theta}_t = \beta \tilde{\theta}_{t-1} + \frac{R_t}{B_t \sigma^2 + R_t} (\hat{\theta}_t - \beta \tilde{\theta}_{t-1}) \quad (5.1)$$

e

$$\Sigma_t = R_t - \frac{R_t^2}{B_t \sigma^2 + R_t} \quad (5.2)$$

Neste caso, $\tilde{\theta}_t$ é o Previsor de Kalman, que pode ser reescrito da seguinte maneira:

$$\tilde{\theta}_t = (1 - q_t) \beta \tilde{\theta}_{t-1} + q_t \hat{\theta}_t,$$

onde

$$q_t = \left(1 + \frac{B_t \sigma^2}{R_t}\right)^{-1}$$

Lembrando que R_t depende de Σ_t , basta dar um "chute" inicial para $\tilde{\theta}_0$ e Σ_0 , para se poder calcular $\tilde{\theta}_t$ de maneira recursiva.

Considerando o caso em que $V \rightarrow \infty$, o que significa que não há associação entre θ_t e θ_{t-1} , $q_t \rightarrow 1$, e o Previsor de Kalman se reduz a $\hat{\theta}_t$, que é a combinação linear das médias amostrais (para o caso 1) e dos estimadores expansão (para o caso 2) dos instantes 0 a t.

Exemplos

1. Suponha que se queira estimar a média populacional no instante t. Neste caso,

$$\theta(t) = \frac{1}{N} \sum_{j=1}^t y_j \quad \text{e} \quad \ell_j = \begin{cases} 1, & \text{se } j=t \\ 0, & \text{se } j=1,2,\dots,t-1 \end{cases}$$

$$B_t = \frac{(1-f)}{n} \quad \text{constante} \quad (f = \frac{n}{N})$$

Suponha que R_t também seja constante ($R_t = R$):

$$q = (1 + \frac{B\sigma^2}{R})^{-1} \implies R = \frac{qB\sigma^2}{1-q}.$$

Então,

$$\begin{aligned} \text{Var}[\theta_t | \mathcal{D}_t] &= R - \frac{R^2}{B\sigma^2 + R} = \frac{qB\sigma^2}{1-q} - \frac{(qB\sigma^2)^2}{(1-q)^2 (B\sigma^2 + \frac{qB\sigma^2}{1-q})} = \\ &= \frac{qB\sigma^2}{1-q} - \frac{q^2 B\sigma^2}{(1-q)^2 + (1-q)q} = \frac{qB\sigma^2 (1-q)}{1-q} = qB\sigma^2 = \\ &= q \left(\frac{1-f}{n} \right) \sigma^2 < \frac{(1-f)}{n} \sigma^2 \quad \text{pois} \quad 0 < q < 1, \end{aligned}$$

o que mostra que o Previsor de Kalman "melhora" a média amostral de uma amostra aleatória simples.

2. Se há interesse em estimar a diferença entre as médias populacionais nos instantes t e $t-1$, ou seja, $\theta_t - \theta_{t-1}$, basta considerar:

$$l_j = \begin{cases} 0 & , \text{ se } j = 1, 2, \dots, t-2 \\ -1 & , \text{ se } j = t-1 \\ 1 & , \text{ se } j = t \end{cases}$$

$$B_t = \frac{2(1-f)}{n} - \frac{2\rho}{n} (\lambda_{(t-1)t} - f)$$

Supondo que B_t e R_t sejam constantes,

$$\text{Var}[\theta_t | \mathcal{D}_t] = R - \frac{R^2}{B\sigma^2 + R}$$

Para minimizar essa variância, é necessário maximizar $B = \frac{2(1-f)}{n} - \frac{2\rho}{n}(\lambda-f)$, o que equivale a escolher $\lambda=1$, ou seja, este procedimento sugere um painel completo como o tipo de amostra a ser utilizado. Essa mesma conclusão foi apresentada por Cochran (1963) e por Duncan/Kalton (1987).

3. Para estimar a média das médias populacionais, basta considerar $\ell_j = \frac{1}{t}$, $j=1,2,\dots,t$.

Supondo que λ_{ij} seja constante, $\lambda_{ij} = \lambda$,
 $\forall i,j \in \{1,2,\dots,t\}$,

$$B_t = \frac{(1-f)}{nt} + \frac{2\rho}{n}(\lambda-f) \frac{1}{t^2} \frac{(t-1)t}{2} = \frac{1-f}{nt} + \frac{\rho(\lambda-f)(t-1)}{nt}$$

Neste caso, $\lambda = 0$, que corresponde à amostras independentes, é o valor que minimiza a $\text{Var}[\theta_t | \mathcal{D}_t]$. Esta conclusão também foi apresentada por Cochran (1963) e Duncan/Kalton (1987).

Observação: utilizando a teoria de Duncan/Horn (1972), é possível obter os mesmos resultados desta seção, sem a suposição de normalidade. Para isto, exige-se somente a existência dos momentos da distribuição envolvidos.

5.3. Modelo de Filtro de Kalman com variâncias desconhecidas

O Modelo de Filtro de Kalman considerado será o mesmo da seção anterior, ou seja,

$$(1) \quad \hat{\theta}_t = \theta_t + e_t \quad , \quad e_t \sim N(0, B_t \sigma^2)$$

$$(2) \quad \theta_t = \beta \theta_{t-1} + \varepsilon_t \quad , \quad \varepsilon_t \sim N(0, V) \quad ,$$

só que agora, $B_t \sigma^2$ e V são ambas desconhecidas.

Sejam σ_1^2 e σ_2^2 as precisões correspondentes a essas variâncias, isto é, $\sigma_1^2 = \frac{1}{B_t \sigma^2}$ e $\sigma_2^2 = \frac{1}{V}$.

Como σ_1^2 e σ_2^2 são desconhecidas, e o modelo é o de Superpopulação Normal, a função de verossimilhança, segundo Broemeling/Diaz/Yusoff (1985), é dada por:

$$\begin{aligned} P(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_t, \theta_1, \theta_2, \dots, \theta_t | \sigma_1^2, \sigma_2^2, \beta) &\propto \\ &\propto (\sigma_1^2)^{t/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^t (\hat{\theta}_i - \theta_i)^2 \sigma_1^2\right\} (\sigma_2^2)^{t/2} \cdot \\ &\cdot \exp\left\{-\frac{1}{2} \sum_{i=1}^t (\theta_i - \beta \theta_{i-1})^2 \sigma_2^2\right\} \quad . \end{aligned}$$

Quanto às precisões σ_1^2 e σ_2^2 , será feita a suposição de que suas prioris sejam Wishart (Johnson/Kotz (1972)).

Então,

$$P(\sigma_1^2) \propto (\sigma_1^2)^{\frac{v_1-2}{2}} \exp\left\{-\frac{\sigma_1^2}{2b_1}\right\}$$

e

$$P(\sigma_2^2) \propto (\sigma_2^2)^{\frac{v_2-2}{2}} \exp\left\{-\frac{\sigma_2^2}{2b_2}\right\} ,$$

onde $v_1 > 0$, $v_2 > 0$, $b_1 > 0$ e $b_2 > 0$ são conhecidos.

A distribuição a posteriori será dada por:

$$\begin{aligned} P(\theta_1, \theta_2, \dots, \theta_t, \sigma_1^2, \sigma_2^2 | \hat{\theta}_1, \dots, \hat{\theta}_t) &\propto \\ &\propto P(\hat{\theta}_1, \dots, \hat{\theta}_t, \theta_1, \dots, \theta_t | \sigma_1^2, \sigma_2^2) P(\sigma_1^2) P(\sigma_2^2) \propto \\ &\propto (\sigma_1^2)^{\frac{v_1+t-2}{2}} \exp\left\{-\frac{1}{2} \left(\sum_{i=1}^t (\hat{\theta}_i - \theta_i)^2 + \frac{1}{b_1} \right) \sigma_1^2 \right\} \cdot \\ &\cdot (\sigma_2^2)^{\frac{v_2+t-2}{2}} \exp\left\{-\frac{1}{2} \left(\sum_{i=1}^t (\theta_i - \beta \theta_{i-1})^2 + \frac{1}{b_2} \right) \sigma_2^2 \right\} . \end{aligned}$$

Lembrando que o objetivo é estimar θ_t , é necessário eliminar σ_1^2 e σ_2^2 e, para isso, integrar a distribuição a posteriori, com relação a esses dois parâmetros. Então,

$$P(\theta_1, \dots, \theta_t | \hat{\theta}_1, \dots, \hat{\theta}_t) = \int_0^\infty \int_0^\infty P(\theta_1, \dots, \theta_t, \sigma_1^2, \sigma_2^2 | \hat{\theta}_1, \dots, \hat{\theta}_t) d\sigma_1^2 d\sigma_2^2 .$$

Escrevendo σ_1^2 e σ_2^2 da maneira original, tem-se que:

$$\begin{aligned}
 P(\theta_1, \dots, \theta_t | \hat{\theta}_1, \dots, \hat{\theta}_t) &\propto \int_0^{\frac{v_1+t-2}{2}} \left(\frac{1}{B_t \sigma^2} \right)^{\frac{v_1+t-2}{2}} \exp \left\{ -\frac{1}{2} \left(\sum_{i=1}^t (\hat{\theta}_i - \theta_i)^2 + \frac{1}{b_1} \right) \frac{1}{B_t \sigma^2} \right\} \frac{(-1)}{(\sigma^2)^2} d\sigma^2 \\
 &\cdot \int_0^{\frac{v_2+t-2}{2}} \left(\frac{1}{V} \right)^{\frac{v_2+t-2}{2}} \exp \left\{ -\frac{1}{2} \left(\sum_{i=1}^t (\theta_i - \beta \theta_{i-1})^2 + \frac{1}{b_2} \right) \frac{1}{V} \right\} \frac{(-1)}{V} dV \propto \\
 &\propto \int_0^{\infty} \left(\sigma^2 \right)^{-\left(\frac{v_1+t+2}{2}\right)} \exp \left\{ -\frac{1}{2B_t} \left(\sum_{i=1}^t (\hat{\theta}_i - \theta_i)^2 + \frac{1}{b_1} \right) \frac{1}{\sigma^2} \right\} d\sigma^2 \cdot \\
 &\cdot \int_0^{\infty} \left(V \right)^{-\left(\frac{v_2+t+2}{2}\right)} \exp \left\{ -\frac{1}{2} \left(\sum_{i=1}^t (\theta_i - \beta \theta_{i-1})^2 + \frac{1}{b_2} \right) \frac{1}{V} \right\} dV
 \end{aligned}$$

Usando o resultado 7 do apêndice, e identificando, para a 1ª integral:

$$q = \frac{v_1+t+2}{2} \quad \text{e} \quad Q = \frac{1}{2B_t} \left(\sum_{i=1}^t (\hat{\theta}_i - \theta_i)^2 + \frac{1}{b_1} \right)$$

e para a 2ª integral:

$$q = \frac{v_2+t+2}{2} \quad \text{e} \quad Q = \frac{1}{2} \left(\sum_{i=1}^t (\theta_i - \beta \theta_{i-1})^2 + \frac{1}{b_2} \right)$$

tem-se que:

$$P(\theta_1, \dots, \theta_t | \hat{\theta}_1, \dots, \hat{\theta}_t) \propto \left(\sum_{i=1}^t (\hat{\theta}_i - \theta_i)^2 + \frac{1}{b_1} \right)^{-\left(\frac{v_1+t}{2}\right)} \left(\sum_{i=1}^t (\theta_i - \beta \theta_{i-1})^2 + \frac{1}{b_2} \right)^{-\left(\frac{v_2+t}{2}\right)}.$$

Conforme sugerido por Broemeling/Diaz/Yusoff (1985), o estimador proposto para θ_t é

$$E[\theta_t | \theta_1, \dots, \theta_{t-1}, \hat{\theta}_1, \dots, \hat{\theta}_t].$$

Mas,

$$\begin{aligned}
 P(\Theta_t | \theta_1, \dots, \theta_{t-1}, \hat{\theta}_1, \dots, \hat{\theta}_t) &\propto \left(\sum_{i=1}^t (\hat{\theta}_i - \theta_i)^2 + \frac{1}{b_1} \right)^{-\left(\frac{\nu_1+t}{2}\right)} \left(\sum_{i=1}^t (\theta_i - \beta\theta_{i-1})^2 + \frac{1}{b_2} \right)^{-\left(\frac{\nu_2+t}{2}\right)} = \\
 &= \left((\hat{\theta}_t - \theta_t)^2 + \sum_{i=1}^{t-1} (\hat{\theta}_i - \theta_i)^2 + \frac{1}{b_1} \right)^{-\left(\frac{\nu_1+t}{2}\right)} \cdot \\
 &\cdot \left((\theta_t - \beta\theta_{t-1})^2 + \sum_{i=1}^{t-1} (\theta_i - \beta\theta_{i-1})^2 + \frac{1}{b_2} \right)^{-\left(\frac{\nu_2+t}{2}\right)}.
 \end{aligned}$$

Usando a seguinte notação:

$$S_1(t-1) = \begin{cases} \frac{1}{b_1} & , \text{ se } t = 1 \\ \sum_{i=1}^{t-1} (\hat{\theta}_i - \theta_i)^2 + \frac{1}{b_1} & , \text{ se } t \neq 1 \end{cases}$$

e

$$S_2(t-1) = \begin{cases} \frac{1}{b_2} & , \text{ se } t = 1 \\ \sum_{i=1}^{t-1} (\theta_i - \beta\theta_{i-1})^2 + \frac{1}{b_2} & , \text{ se } t \neq 1 \end{cases}$$

$$\begin{aligned}
 P(\Theta_t | \theta_1, \dots, \theta_{t-1}, \hat{\theta}_1, \dots, \hat{\theta}_t) &\propto \left((\hat{\theta}_t - \theta_t)^2 + S_1(t-1) \right)^{-\left(\frac{\nu_1+t}{2}\right)} \left((\theta_t - \beta\theta_{t-1})^2 + S_2(t-1) \right)^{-\left(\frac{\nu_2+t}{2}\right)} = \\
 &= \left[S_1(t-1) \right]^{-\left(\frac{\nu_1+t}{2}\right)} \left[1 + S_1^{-1}(t-1) (\hat{\theta}_t - \theta_t)^2 \right]^{-\left(\frac{\nu_1+t}{2}\right)} \cdot \\
 &\cdot \left[S_2(t-1) \right]^{-\left(\frac{\nu_2+t}{2}\right)} \left[1 + S_2^{-1}(t-1) (\theta_t - \beta\theta_{t-1})^2 \right]^{-\left(\frac{\nu_2+t}{2}\right)} \propto \\
 &\propto \left[1 + S_1^{-1}(t-1) (\hat{\theta}_t - \theta_t)^2 \right]^{-\left(\frac{\nu_1+t}{2}\right)} \left[1 + S_2^{-1}(t-1) (\theta_t - \beta\theta_{t-1})^2 \right]^{-\left(\frac{\nu_2+t}{2}\right)} = \\
 &= T
 \end{aligned}$$

De acordo com Broemeling/Diaz/Yusoff (1985), essa densidade pode ser aproximada por uma densidade Normal, cuja esperança é o estimador procurado. Para realizar essa aproximação, foram seguidos os passos de Press (1972).

Em primeiro lugar foi usado o seguinte fato:

$T = \exp. \ln(T)$. Então,

$$T = \exp \left\{ -\frac{1}{2} \left[(v_1+t) \ln(1+S_1^{-1}(t-1) (\hat{\theta}_t - \theta_t)^2) + (v_2+t) \ln(1+S_2^{-1}(t-1) (\theta_t - \beta \theta_{t-1})^2) \right] \right\}$$

Mas, $\ln(1+x) \approx x$, para $-1 < x < 1$. Neste caso, para

$S_1^{-1}(t-1) (\hat{\theta}_t - \theta_t)^2$ e $S_2^{-1}(t-1) (\theta_t - \beta \theta_{t-1})^2$ pequenos,

$$\begin{aligned} T &\approx \exp \left\{ -\frac{1}{2} \left[(v_1+t) S_1^{-1}(t-1) (\hat{\theta}_t - \theta_t)^2 + (v_2+t) S_2^{-1}(t-1) (\theta_t - \beta \theta_{t-1})^2 \right] \right\} = \\ &= \exp \left\{ -\frac{1}{2} \left[\frac{(v_1+t) (\hat{\theta}_t^2 - 2\hat{\theta}_t \theta_t + \theta_t^2)}{S_1(t-1)} + \frac{(v_2+t) (\theta_t^2 - 2\beta \theta_t \theta_{t-1} + \beta^2 \theta_{t-1}^2)}{S_2(t-1)} \right] \right\} = \\ &= \exp \left\{ -\frac{1}{2} \left[\left(\frac{v_1+t}{S_1(t-1)} + \frac{v_2+t}{S_2(t-1)} \right) \theta_t^2 - 2\theta_t \left(\frac{v_1+t}{S_1(t-1)} \hat{\theta}_t + \frac{v_2+t}{S_2(t-1)} \beta \theta_{t-1} \right) + \right. \right. \\ &\quad \left. \left. + \frac{v_1+t}{S_1(t-1)} \hat{\theta}_t^2 + \frac{v_2+t}{S_2(t-1)} \beta^2 \theta_{t-1}^2 \right] \right\}. \end{aligned}$$

Chamando $a = \frac{v_1+t}{S_1(t-1)}$ e $b = \frac{v_2+t}{S_2(t-1)}$,

$$T \approx \exp \left\{ -\frac{1}{2} \left[\frac{\theta_t^2}{\frac{1}{a+b}} - \frac{2\theta_t (a\hat{\theta}_t + b\beta\theta_{t-1})}{\frac{1}{a+b}} \frac{1}{a+b} + \frac{a\hat{\theta}_t^2 + b\beta^2\theta_{t-1}^2}{\frac{1}{a+b}} \frac{1}{a+b} \right] \right\}$$

Completando o quadrado,

$$T \cong \exp \left\{ -\frac{1}{2}(a+b) \left(\theta_t^2 - 2\theta_t \left(\frac{a\hat{\theta}_t + b\beta\theta_{t-1}}{a+b} \right) + \left(\frac{a\hat{\theta}_t + b\beta\theta_{t-1}}{a+b} \right)^2 + \frac{-2ab\beta\hat{\theta}_t\theta_{t-1} + ab\beta^2\theta_{t-1}^2 + ab\theta_t^2}{(a+b)^2} \right) \right\} \propto$$

$$\propto \exp \left\{ -\frac{1}{2}(a+b) \left(\theta_t^2 - 2\theta_t \left(\frac{a\hat{\theta}_t + b\beta\theta_{t-1}}{a+b} \right) + \left(\frac{a\hat{\theta}_t + b\beta\theta_{t-1}}{a+b} \right)^2 \right) \right\}$$

Logo, T corresponde aproximadamente a uma Normal de média $\frac{a}{a+b} \hat{\theta}_t + \frac{b}{a+b} \beta\theta_{t-1}$ e variância $\frac{1}{a+b}$.

O estimador proposto, definido de maneira recursiva, será, portanto,

$$\hat{\theta}_t = q_t \hat{\theta}_t + (1-q_t) \tilde{\beta}\theta_{t-1}, \quad (5.3)$$

onde

$$q_t = \frac{a}{a+b} = \frac{(v_1+t)S_2(t-1)}{(v_1+t)S_2(t-1) + (v_2+t)S_1(t-1)}, \quad (5.4)$$

$$S_1(t-1) = \begin{cases} \frac{1}{b_1}, & \text{se } t = 1 \\ \sum_{i=1}^{t-1} (\hat{\theta}_i - \tilde{\theta}_i)^2 + \frac{1}{b_1}, & \text{se } t \neq 1 \end{cases} \quad (5.5)$$

$$e \quad S_2(t-1) = \begin{cases} \frac{1}{b_2}, & \text{se } t = 1 \\ \sum_{i=1}^{t-1} (\tilde{\theta}_i - \beta\hat{\theta}_{i-1})^2 + \frac{1}{b_2}, & \text{se } t \neq 1, \end{cases} \quad (5.6)$$

ou seja, é uma média ponderada de dois estimadores, $\hat{\theta}(t)$ e $\tilde{\beta}\theta(t-1)$.

CAPÍTULO 6

OBTENÇÃO DE ESTIMADORES ÓTIMOS ATRAVÉS DA TEORIA DE MÍNIMOS QUADRADOS

6.1. Introdução

O objetivo, neste capítulo, é apresentar um estimador para θ , derivado através da Teoria de Mínimos Quadrados, quando os parâmetros são variáveis aleatórias.

Este procedimento foi utilizado por Jones (1980), baseado nas técnicas apresentadas por Rao (1973), pág. 234.

Do mesmo modo que no capítulo anterior, será considerado somente o caso em que $F(t)$ está associada a uma variável discreta e, portanto, $\theta = \sum_{j=0}^T c(j)\theta(j)p_j$. Além disso, conforme já explicitado anteriormente, quando $c(j)$ e p_j são especificados, pode-se escrever:

$$l_j = c(j)p_j \quad \text{e} \quad \theta_t = \sum_{j=1}^t l_j \theta(j) ,$$

ou seja, θ_t é uma combinação linear dos $\theta(j)$'s.

O critério de otimalidade a ser utilizado será o seguinte:

$\hat{\theta}_t^*$ é ótimo para θ_t se as duas condições abaixo estiverem satisfeitas:

$$(1) \quad E[\hat{\theta}_t^* - \theta_t] = 0$$

$$(2) \quad E[\hat{\theta}_t^* - \theta_t]^2 \leq E[\hat{\theta}_t - \theta_t]^2, \quad \forall \hat{\theta}_t \text{ satisfazendo (1).}$$

Também, como no Capítulo 5, serão apresentados resultados para 2 casos: quando $\theta(t)$ é a média populacional ou o total populacional no instante t .

Novamente, serão feitas as seguintes suposições:

$$E[\underline{Y}(t)] = \mu \underline{1}_N, \quad V[\underline{Y}(t)] = \sigma^2 \underline{I}$$

e

$$\text{Cov}[Y_k(t), Y_{k'}(t')] = \begin{cases} \sigma^2 \rho & , \quad k = k' , \quad t \neq t' \\ 0 & , \quad k \neq k' , \quad t \neq t' \end{cases} \quad (6.1)$$

onde σ^2 e ρ são conhecidos. O significado dessas suposições é que não há correlação entre observações associadas a diferentes elementos, k e $k' \in P_1(t)$, para qualquer instante e que, a correlação entre observações correspondentes a um mesmo elemento k , para instantes diferentes, é ρ .

É importante ressaltar que, neste caso, a suposição de normalidade não é necessária e portanto, este método se apresenta como uma alternativa ao de Filtro de Kalman, quando a distribuição de $\underline{Y}(t)$ é desconhecida.

Outra vantagem dessa abordagem é que a equação do sistema, que dá a forma da relação entre valores da população para instantes sucessivos, é ignorada.

Então, de acordo com as suposições feitas acima, para um determinado instante j e um determinado elemento k , pode-se escrever $Y_k(j)$ da seguinte forma:

$$Y_k(j) = \mu + e_k(j), \quad (6.2)$$

onde $E[e_k(j)] = 0$ e $\text{Var}[e_k(j)] = \sigma^2$, ou, considerando-se os N elementos da população no instante j ,

$$\underline{Y}(j) = \mu \underline{1}_N + \underline{e}(j) \quad (6.3)$$

onde $\underline{Y}'(j) = (Y_1(j), Y_2(j), \dots, Y_N(j))$,

$\underline{e}(j)$ é o vetor dos erros amostrais, ou seja,

$$\underline{e}'(j) = (e_1(j), e_2(j), \dots, e_N(j))$$

$$E[\underline{e}(j)] = \underline{0} \quad \text{e} \quad V[\underline{e}(j)] = \sigma^2 \underline{I}$$

Quanto a μ , duas alternativas serão consideradas:
 μ conhecida e μ desconhecida.

6.2. Obtenção dos estimadores, considerando μ conhecida.

Uma amostra de n elementos é selecionada da população, para cada instante e λ_{ij} denota a proporção, na amostra, de elementos comuns aos instantes i e j ($\lambda_{ij} = \frac{m_{ij}}{n}$).

Caso 1. $\theta(j)$ é a média populacional no instante j ,

$$\theta(j) = \frac{1}{N} \underline{1}'_N \underline{Y}(j)$$

i) multiplicando-se a equação (6.3) por $\frac{1}{N} \underline{1}'_N$,

$$\frac{1}{N} \underline{1}'_N \underline{Y}(j) = \frac{1}{N} \underline{1}'_N \mu \underline{1}_N + \frac{1}{N} \underline{1}'_N \underline{e}(j)$$

ou seja,

$$\theta(j) = \mu + \frac{1}{N} \mathbf{1}'_{\sim N} e(j) \quad , \quad (6.4)$$

onde

$$E[\frac{1}{N} \mathbf{1}'_{\sim N} e(j)] = 0 \quad \text{e} \quad \text{Var}[\frac{1}{N} \mathbf{1}'_{\sim N} e(j)] = \frac{\sigma^2}{N} .$$

ii) multiplicando-se a equação (6.3) por $\frac{1}{n} \mathbf{I}'_{\sim j}$, onde $\mathbf{I}'_{\sim j}$ é o vetor (de tamanho $1 \times N$) de 0's e 1's: 1 se o elemento correspondente àquela posição pertence à amostra no instante j e zero caso contrário,

$$\frac{1}{n} \mathbf{I}'_{\sim j} Y(j) = \frac{1}{n} \mathbf{I}'_{\sim j} \mu \mathbf{1}_{\sim N} + \frac{1}{n} \mathbf{I}'_{\sim j} e(j) \quad ,$$

ou

$$\hat{\theta}(j) = \mu + \frac{1}{n} \mathbf{I}'_{\sim j} e(j) \quad , \quad (6.5)$$

onde $\hat{\theta}(j)$ é a média amostral no instante j ,

$$E[\frac{1}{n} \mathbf{I}'_{\sim j} e(j)] = 0 \quad \text{e} \quad \text{Var}[\frac{1}{n} \mathbf{I}'_{\sim j} e(j)] = \frac{\sigma^2}{n} .$$

Agora, pode-se extrair o valor de μ da equação (6.4),

ou seja,

$$\mu = \theta(j) - \frac{1}{N} \mathbf{1}'_{\sim N} e(j) \quad ,$$

e substituí-lo na equação (6.5). Daí, tem-se que:

$$\hat{\theta}(j) = \theta(j) + \varepsilon(j) \quad , \quad (6.6)$$

onde

$$\varepsilon(j) = \frac{1}{n} \mathbf{I}'_{\sim j} e(j) - \frac{1}{N} \mathbf{1}'_{\sim N} e(j) \quad ,$$

$$E[\varepsilon(j)] = 0 \quad \text{e} \quad \text{Var}[\varepsilon(j)] = (\frac{1}{n} - \frac{1}{N}) \sigma^2 ,$$

pois

$$E[\varepsilon(j)] = E[(\frac{1}{n} \mathbf{I}'_{\sim j} - \frac{1}{N} \mathbf{1}'_{\sim N}) e(j)] = (\frac{1}{n} \mathbf{I}'_{\sim j} - \frac{1}{N} \mathbf{1}'_{\sim N}) E[e(j)] = 0$$

e

$$\begin{aligned} \text{Var}[\varepsilon(j)] &= \text{Var}\left[\left(\frac{1}{n}I_j' - \frac{1}{N}1_N'\right)e(j)\right] = \left(\frac{1}{n}I_j' - \frac{1}{N}1_N'\right)V[e(j)]\left(\frac{1}{n}I_j - \frac{1}{N}1_N\right) = \\ &= \left(\frac{1}{n}I_j' - \frac{1}{N}1_N'\right)\sigma^2 I\left(\frac{1}{n}I_j - \frac{1}{N}1_N\right) = \\ &= \left[n\left(\frac{1}{n} - \frac{1}{N}\right)^2 + (N-n)\frac{1}{N}\right]\sigma^2 = \left[\frac{(N-n)^2}{Nn} + \frac{N-n}{N}\right]\sigma^2 = \\ &= \frac{N^2 - 2Nn + n^2 + Nn - n^2}{N^2 n} \sigma^2 = \frac{N-n}{Nn} \sigma^2 = \left(\frac{1}{n} - \frac{1}{N}\right)\sigma^2 = \frac{N-n}{n} \frac{\sigma^2}{N} . \end{aligned}$$

A equação (6.6) equivale à equação das observações no modelo de Filtro de Kalman ($\theta_t = \hat{\theta}_t + e_t$).

Até agora foram utilizadas somente as observações correspondentes ao instante j e portanto, a correlação que existe entre observações associadas à mesma unidade, para instantes diferentes, ainda não foi considerada.

Então, seja $\tilde{\theta}(t)$ o vetor dos $\theta(j)$'s para os instantes $1, 2, \dots, t$, isto é,

$$\tilde{\theta}'(t) = (\theta(1), \theta(2), \dots, \theta(t)) .$$

Analogamente,

$$\hat{\tilde{\theta}}'(t) = (\hat{\theta}(1), \hat{\theta}(2), \dots, \hat{\theta}(t)) \quad \text{e} \quad \tilde{\varepsilon}'(t) = (\varepsilon(1), \varepsilon(2), \dots, \varepsilon(t)) .$$

Daí, pode-se escrever:

$$\hat{\tilde{\theta}}(t) = \tilde{\theta}(t) + \tilde{\varepsilon}(t) , \quad (6.7)$$

onde

$$E[\tilde{\varepsilon}(t)] = \tilde{0} \quad \text{e} \quad \text{Var}[\tilde{\varepsilon}(t)] = \tilde{\Sigma} = \frac{\sigma^2}{Nn}[a_{ij}]$$

com

$$a_{ij} = \begin{cases} (N-n) & , \text{ se } i = j \\ (N\lambda_{ij} - n)\rho & , \text{ se } i \neq j \end{cases} ,$$

$$\begin{aligned}
 & \text{pois} \quad \text{Cov}[\varepsilon(i), \varepsilon(j)] = \\
 & = \text{Cov}\left[\frac{1}{n} \sum_{k \in S_i} e_k(i) - \frac{1}{N} \sum_{k=1}^N e_k(i), \frac{1}{n} \sum_{k \in S_j} e_k(j) - \frac{1}{N} \sum_{k=1}^N e_k(j)\right] = \\
 & = \frac{1}{n^2} m_{ij} \sigma^2 \rho - \frac{1}{Nn} n \sigma^2 \rho - \frac{1}{Nn} n \sigma^2 \rho + \frac{1}{N^2} N \sigma^2 \rho = \\
 & = \left(\frac{\lambda_{ij}}{n} - \frac{1}{N}\right) \sigma^2 \rho = \frac{N\lambda_{ij} - n}{Nn} \sigma^2 \rho,
 \end{aligned}$$

onde S_j representa a amostra no instante j e m_{ij} é o número de elementos comuns às amostras S_i e S_j .

Por outro lado, $\tilde{\theta}(t)$ é um vetor de variáveis aleatórias, tais que:

$$E[\tilde{\theta}(t)] = \mu \underline{1}_N \quad \text{e} \quad V[\tilde{\theta}(t)] = \underline{\tau} = \frac{\sigma^2}{N} [a_{ij}],$$

$$\text{com} \quad a_{ij} = \begin{cases} 1, & \text{se } i=j, \\ \rho, & \text{se } i \neq j \end{cases},$$

pois

$$\begin{aligned}
 \text{Cov}[\theta(i), \theta(j)] &= \text{Cov}\left[\frac{1}{N} \sum_{k=1}^N y_k(i), \frac{1}{N} \sum_{k=1}^N y_k(j)\right] = \\
 &= \frac{1}{N^2} N \sigma^2 \rho = \frac{\sigma^2 \rho}{N}.
 \end{aligned}$$

Identificando essas matrizes com as do resultado 8 do apêndice e considerando a seguinte combinação linear $\underline{\ell}' = (\ell_1, \ell_2, \dots, \ell_t)$, tem-se que o estimador de $\theta_t = \tilde{\theta}'(t)$ é dado por:

$$\hat{\theta}_t = \underline{\mu}' \underline{\ell} - \underline{\mu}' \underline{\Sigma}^{-1} (\underline{\tau}^{-1} + \underline{\Sigma}^{-1})^{-1} \underline{\ell} + \underline{\ell}' (\underline{\tau}^{-1} + \underline{\Sigma}^{-1})^{-1} \underline{\Sigma}^{-1} \hat{\theta}(t)$$

e seu erro quadrático médio é

$$E[\theta_t - \hat{\theta}_t]^2 = \underline{\ell}' (\underline{\tau}^{-1} + \underline{\Sigma}^{-1})^{-1} \underline{\ell}$$

Se $\lambda_{ij} = \lambda$, para $\forall i, \forall j$, isto é, a proporção mantida na amostra entre dois instantes diferentes é constante, a matriz $\underline{\Sigma}$ tem a seguinte forma:

$$\begin{pmatrix} a & b & b & \dots & b \\ b & a & b & \dots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b & b & b & \dots & a \end{pmatrix} \quad (6.8)$$

Isto significa que $\text{Var}[\varepsilon(j)] = (\frac{1}{n} - \frac{1}{N}) \sigma^2$, $\forall j$, constante para todos os instantes e

$$\text{Cov}[\varepsilon(i), \varepsilon(j)] = \frac{(N\lambda - n)}{Nn} \sigma^2 \rho, \quad \forall i, \forall j,$$

também constantes.

Neste caso, pode-se usar o resultado 4 do apêndice para se inverter as matrizes $\underline{\Sigma}$ e $\underline{\tau}$ (que também tem a mesma forma).

O estimador $\hat{\theta}_t$ pode, então, ser escrito como:

$$\hat{\theta}_t = A \mu \sum_{j=1}^t \ell_j + B \sum_{j=1}^t \ell_j \hat{\theta}(j) + (t-1) C \sum_{j=1}^t \hat{\theta}(j) - \frac{1}{(t-1)} \sum_{\substack{i=1 \\ i \neq j}}^t \ell_i \quad (6.9)$$

onde

$$A = 1 - \frac{n}{N} \frac{(1+(t-1)\rho)}{(1+(t-1)\lambda\rho)}, \quad (6.10)$$

$$B = \frac{n}{N} \frac{1+(t-2)\lambda\rho - (t-1)\lambda\rho^2}{(1+(t-1)\lambda\rho)(1-\lambda\rho)} \quad (6.11)$$

e

$$C = \frac{n}{N} \rho \frac{1-\lambda}{(1+(t-1)\lambda\rho)(1-\lambda\rho)} \quad (6.12)$$

O resultado 9 do apêndice apresenta a derivação dessas fórmulas.

A relação que existe entre esses coeficientes é $A + B + (t-1)C = 1$ e, então, $\hat{\theta}_t$ pode ser, convenientemente, escrito da seguinte forma:

$$\hat{\theta}_t = B \sum_{j=1}^t \lambda_j \hat{\theta}(j) + (1-B) \hat{\theta}_t ,$$

ou seja, $\hat{\theta}_t$ é a média ponderada de dois estimadores: um que é a combinação linear das médias amostrais e outro, $\hat{\theta}_t$, que por sua vez também é uma média ponderada, dado por:

$$\hat{\theta}_t = \frac{A}{1-B} \mu \sum_{j=1}^t \lambda_j + (1 - \frac{A}{1-B}) \sum_{j=1}^t \hat{\theta}(j) \frac{1}{(t-1)} \sum_{\substack{i=1 \\ i \neq j}}^t 1 \quad (6.13)$$

Caso 2: $\theta(j)$ é o total populacional no instante j ,

$$\theta(j) = \mathbf{1}'_{\tilde{N}} \mathbf{Y}(j)$$

Neste caso, a equação (6.3) será multiplicada por $\mathbf{1}'_{\tilde{N}}$ e por $\frac{N}{n} \mathbf{I}'_{\tilde{j}}$, isto é, $\hat{\theta}(j)$ é o estimador expansão. Então,

$$\theta(j) = N\mu + \mathbf{1}'_{\tilde{N}} \mathbf{e}(j) , \quad (6.14)$$

onde

$$E[\mathbf{1}'_{\tilde{N}} \mathbf{e}(j)] = 0 \quad \text{e} \quad \text{Var}[\mathbf{1}'_{\tilde{N}} \mathbf{e}(j)] = N\sigma^2.$$

e

$$\hat{\theta}(j) = N\mu + \frac{N}{n} \frac{1}{j} e(j) , \quad (6.15)$$

onde

$$E\left[\frac{N}{n} \frac{1}{j} e(j)\right] = 0 \quad \text{e} \quad \text{Var}\left[\frac{N}{n} \frac{1}{j} e(j)\right] = \frac{N^2 \sigma^2}{n}$$

Analogamente ao caso da média, extraíndo-se o valor de $N\mu$ da equação (6.14) e substituindo-o na equação (6.15), obtem-se:

$$\hat{\theta}(j) = \theta(j) + \varepsilon(j) , \quad (6.16)$$

onde $\varepsilon(j) = \frac{N}{n} \frac{1}{j} e(j) - \frac{1}{N} e(j) , \quad E[\varepsilon(j)] = 0$

e

$$\text{Var}[\varepsilon(j)] = N^2 \left(\frac{1}{n} - \frac{1}{N} \right) \sigma^2$$

Considerando-se os "t" instantes:

$$\hat{\theta}(t) = \theta(t) + \varepsilon(t) , \quad (6.17)$$

onde $E[\varepsilon(t)] = 0 \quad \text{e} \quad V[\varepsilon(t)] = \frac{N}{n} \sigma^2 [a_{ij}]$

com
$$a_{ij} = \begin{cases} (N-n) & , \text{ se } i=j \\ (N\lambda_{ij}-n)\rho & , \text{ se } i \neq j \end{cases} .$$

Por outro lado, $E[\theta(t)] = N\mu \frac{1}{N}$ e $V[\theta(t)] = N\sigma^2 [a_{ij}]$,

com
$$a_{ij} = \begin{cases} 1 & , \text{ se } i=j \\ \rho & , \text{ se } i \neq j \end{cases} .$$

Comparando-se com o caso anterior, nota-se que as matrizes de dispersão do total são as mesmas que da média (Σ e τ), multiplicadas por N^2 .

Usando-se o mesmo procedimento,

$$\hat{\theta}_t = N\mu' \tilde{\ell} - N\mu' \tilde{\Sigma}^{-1} (\tilde{\tau}^{-1} + \tilde{\Sigma}^{-1})^{-1} \tilde{\ell} + \tilde{\ell}' (\tilde{\tau}^{-1} + \tilde{\Sigma}^{-1})^{-1} \tilde{\Sigma}^{-1} \hat{\theta}(t), \quad (6.18)$$

onde $\hat{\theta}'(t) = (\hat{\theta}(1), \hat{\theta}(2), \dots, \hat{\theta}(t))$ e $\hat{\theta}(j) = \frac{N}{n} \mathbf{I}'_j \mathbf{Y}(j)$,

$j = 1, 2, \dots, t$ e o erro quadrático médio é o mesmo:

$$\tilde{\ell}' (\tilde{\tau}^{-1} + \tilde{\Sigma}^{-1})^{-1} \tilde{\ell}.$$

Como alternativa, $\hat{\theta}_t$ pode ser escrito como:

$$\hat{\theta}_t = NA\mu \sum_{j=1}^t \ell_j + B \sum_{j=1}^t \ell_j \hat{\theta}(j) + (t-1)C \sum_{j=1}^t \hat{\theta}(j) \frac{1}{(t-1)} \sum_{\substack{i=1 \\ i \neq j}}^t \ell_i \quad (6.19)$$

que é o mesmo estimador da média multiplicado por N.

Caso particular: $\rho = 0$

Analisando o caso particular em que $\rho = 0$, ou seja, não há correlação entre observações correspondentes à mesma unidade para instantes diferentes, tem-se que:

$$A = 1-f, \quad B = f \quad \text{e} \quad C = 0, \quad f = \frac{n}{N}$$

e o estimador $\hat{\theta}_t$ se reduz a

$$\hat{\theta}_t = (1-f)\mu \sum_{j=1}^t \ell_j + f \sum_{j=1}^t \ell_j \hat{\theta}(j),$$

que é uma média ponderada entre duas quantidades, que são a combinação linear dos elementos do vetor μ ($\ell'\mu$) e a combinação linear dos $\hat{\theta}(j)$'s ($\ell'\hat{\theta}(t)$), e o fator de ponderação é a fração amostral f .

Para tentar analisar esse estimador, basta escrever θ_t como a soma de duas partes, uma que depende da amostra e outra que não depende da amostra, isto é,

$$\begin{aligned}\theta_t &= \sum_{j=1}^t \ell_j \theta(j) = \sum_{j=1}^t \ell_j \frac{1}{N} \sum_{k=1}^N Y_k(j) = \\ &= \frac{1}{N} \sum_{j=1}^t \ell_j \left(\sum_{k \in S_j} Y_k(j) + \sum_{k \notin S_j} Y_k(j) \right) = \\ &= \frac{n}{N} \sum_{j=1}^t \ell_j \hat{\theta}(j) + \frac{(N-n)}{N} \sum_{j=1}^t \ell_j \frac{1}{N-n} \sum_{k \notin S_j} Y_k(j).\end{aligned}$$

Comparando-se com o estimador, tem-se que, depois de retirada a amostra, a primeira parcela é conhecida e cada elemento da segunda parcela é estimado por μ .

Ainda dentro desse caso ($\rho=0$), serão consideradas, como exemplo, algumas particulares combinações lineares:

i) estimar a média no instante t

$$\ell_j = \begin{cases} 1, & \text{se } j=t \\ 0, & \text{se } j = 1, 2, \dots, t-1 \end{cases}$$

$$\hat{\theta}_t = (1-f)\mu + f\hat{\theta}(t)$$

ii) estimar $\theta(t) - \theta(t-1)$

$$\ell_j = \begin{cases} 1, & \text{se } j=t \\ -1, & \text{se } j=t-1 \\ 0, & \text{se } j=1, 2, \dots, t-2 \end{cases}$$

$$\hat{\theta}_t = f[\hat{\theta}(t) - \hat{\theta}(t-1)]$$

iii) estimar a média das médias

$$\ell_j = \frac{1}{t} \quad , \quad j=1,2,\dots,t$$

$$\hat{\theta}_t = (1-f)\mu + \frac{f}{t} \sum_{j=1}^t \hat{\theta}(j)$$

6.3. Obtenção dos estimadores, considerando μ desconhecida.

Aplicando a segunda parte do resultado 8 do apêndice, tem-se que:

$$\hat{\theta}_t = \sum_{j=1}^t \ell_j \hat{\theta}(j) \quad ,$$

onde $\hat{\theta}(j)$ é a média amostral para o caso da média e $\hat{\theta}$ é o estimador expansão para o caso do total.

Quanto ao seu erro quadrático médio, tem-se:

$$E[\theta_t - \hat{\theta}_t]^2 = \sum_{j=1}^t \ell_j^2$$

para a média e

$$E[\theta_t - \hat{\theta}_t]^2 = N^2 \sum_{j=1}^t \ell_j^2$$

para o total.

Se $\lambda_{ij} = \lambda$, constante, então o erro quadrático médio também pode ser escrito como:

$$D((N-n) \sum_{j=1}^t \ell_j^2 + (N\lambda-n) \rho \sum_{j=1}^t \ell_j \sum_{\substack{i=1 \\ i \neq j}}^t \ell_i) \quad ,$$

onde $D = \frac{\sigma^2}{nN}$, para o caso da média e $D = \frac{N\sigma^2}{n}$, para o caso do total.

Para minimizar essa expressão com relação a λ , basta minimizar $(N\lambda - n)\rho \sum_{j=1}^t \ell_j \sum_{\substack{i=1 \\ i \neq j}}^t \ell_i$, que é uma reta. O valor de λ , que torna o erro quadrático médio mínimo, dependerá da combinação linear de interesse e do coeficiente de correlação.

Como a característica medida para os vários instantes de tempo é sempre a mesma, será feita a suposição de que $\rho > 0$ e então:

1) Se $\sum_{j=1}^t \ell_j \sum_{\substack{i=1 \\ i \neq j}}^t \ell_i = 0$, qualquer valor de λ , $0 \leq \lambda \leq 1$, pode ser usado.

Exemplo: estimar a média no instante t : $\theta(t)$

$\hat{\theta}_t = \hat{\theta}(t)$ média amostral no instante t

$$E[\theta_t - \hat{\theta}_t]^2 = \frac{\sigma^2(N-n)}{nN}$$

Qualquer λ , ou seja, qualquer tipo de amostra é equivalente para este caso.

2) Se $\sum_{j=1}^t \ell_j \sum_{\substack{i=1 \\ i \neq j}}^t \ell_i < 0$, a reta tem coeficiente angular negativo e então, o melhor valor de λ é $\lambda=1$, que sugere um painel completo como o tipo de amostra a ser utilizado.

Exemplo: estimar a variação das médias nos instantes t e $t-1$:

$$\theta(t) - \theta(t-1)$$

$\hat{\theta}_t = \hat{\theta}(t) - \hat{\theta}(t-1)$ diferença entre as médias amostrais

$$E[\theta_t - \hat{\theta}_t]^2 = \frac{2\sigma^2}{nN} ((N-n) - (N\lambda-n)\rho)$$

A amostra sugerida é o painel completo, pois o valor de λ que minimiza o erro quadrático médio é $\lambda = 1$. Essa conclusão, apresentada por Cochran (1963) e Duncan/Kalton (1985) já foi citada, anteriormente, neste trabalho na página 83.

3) Se $\sum_{j=1}^t \ell_j \sum_{\substack{i=1 \\ i \neq j}}^t \ell_i > 0$, a reta tem coeficiente angular positivo e então, a sugestão aponta para $\lambda = 0$, que corresponde à amostras independentes.

Exemplo: estimar a média das médias: $\frac{1}{t} \sum_{j=1}^t \theta(j)$

$$\hat{\theta}_t = \frac{1}{t} \sum_{j=1}^t \hat{\theta}(j)$$

$$E[\theta_t - \hat{\theta}_t]^2 = \frac{\sigma^2}{nNt} ((N-n) + (N\lambda-n)\rho(t-1))$$

Neste caso, $\lambda = 0$ é escolhido, o que corresponde à amostras independentes. Como no exemplo anterior, esta conclusão também foi apresentada por Cochran (1963) e Duncan/Kalton (1985) e citada na página 83.

CAPÍTULO 7

ILUSTRAÇÕES NUMÉRICAS

7.1. Introdução

O objetivo, neste capítulo, é ilustrar numericamente, o comportamento dos estimadores (previsores) de θ obtidos nos capítulos 4, 5 e 6. Em todos os exemplos $\theta(t)$ será a média populacional no instante t , ou seja,

$$\theta(t) = \frac{1}{N} \sum_{i=1}^N y_i(t) .$$

As ilustrações dos capítulo 4 e 5 serão realizadas sobre um conjunto de dados reais e do capítulo 6, sobre dados simulados.

7.2. Um exemplo com dados reais

Os dados reais referem-se ao número de empregados de um grupo de indústrias do setor de Papel e Papelão do Estado de São Paulo, no último dia de cada mês do ano de 1986. Esses dados fazem parte da pesquisa "Levantamento de Conjuntura - Índices FIESP", realizada mensalmente pela Federação das Indústrias do Estado de São Paulo.

A fim de melhor entender o tratamento imposto aos dados, algumas explicações, dadas a seguir, sobre esse levanta -

mento, se fazem necessárias.

A FIESP conta com um painel de indústrias, que respondem mensalmente a um questionário. Um dos itens desse questionário é o total de pessoal ocupado no último dia do mês de referência. Com base nesse item, deseja-se estimar o número médio de empregados por indústria, para cada mês.

Alguns desses informantes atrasam o envio do questionário e, como se deseja uma informação recente, a estimativa dessa média é feita através da média dos dados obtidos até determinado dia (nessa data o painel está incompleto). Posteriormente, com o recebimento dos demais questionários, nova estimativa é calculada, agora com todos os informantes do painel, e esse novo valor passa a ser a estimativa definitiva.

Portanto, existe uma população, que é o conjunto de todas as indústrias de Papel e Papelão do Estado de São Paulo; existe um painel, que dará a estimativa definitiva do número médio de pessoas empregadas por indústria e existe um subconjunto do painel, composto pelos informantes que responderam ao questionário até uma data fixada, que dará a estimativa preliminar dessa média.

O objetivo é encontrar uma estimativa preliminar que seja próxima da estimativa definitiva. Por esse motivo, o painel completo será visualizado como sendo a população, e a média desse painel será a função a ser estimada, com base nos dados disponíveis até a data de interesse.

Assim sendo, a população $P_1(t)$ conta com $N = 32$ elementos. Atualmente todas as informações do painel para o ano

de 1986 já estão disponíveis. As médias correspondentes foram calculadas e serão comparadas com o atual método de estimação (média das informações obtidas até a data de fechamento do levantamento) e com os métodos propostos nos Capítulos 4 e 5 deste trabalho.

É importante ressaltar que o método utilizado atualmente considera somente as informações obtidas para o mês de referência, enquanto que os métodos propostos aqui, utilizarão também os dados correspondentes ao mês anterior ao de referência, ou seja, $P_2(T)$ contará com dois instantes, denominados "1" e "2", $P_2(T) = \{1, 2\}$.

A opção por somente dois instantes foi devida à simplicidade da utilização dos novos métodos.

O Estimador Proposto no Capítulo 4

No Capítulo 4, o estimador proposto para este caso é o seguinte:

$$\hat{\theta}^* = \rho K(1-\lambda) (\bar{Y}_u(1) - \bar{Y}_m(1)) + \bar{Y}_m(2) , \quad (7.1)$$

onde ρ é o coeficiente de correlação entre observações correspondentes à mesma unidade nos instantes 1 e 2, K é o quociente entre os desvios padrões da variável em questão, para os instantes 1 e 2, $K = \frac{\sigma_2}{\sigma_1}$ e λ é a proporção de elementos comuns aos dois instantes, com relação à amostra total.

Como ρ e K são desconhecidos, os dados completos de 1985 e de 1986 foram utilizados a fim de estimá-los. O que se observou é que tanto ρ quanto K estão sempre muito próximos do valor 1, que será, portanto, utilizado para efeitos práticos.

Essa suposição parece bastante razoável uma vez que a variável de interesse é o total de pessoal ocupado, que varia muito pouco de um mês para outro e que tem variância praticamente constante ao longo dos meses.

Quanto a λ , os valores utilizados na fórmula (7.1) estão apresentados na tabela 7.1.

Tabela 7.1. - Valores de λ utilizados em (7.1) para os meses de fevereiro/86 a dezembro/86.

Mês	λ
Fevereiro	0,500
Março	0,469
Abril	0,469
Maio	0,688
Junho	0,656
Julho	0,656
Agosto	0,656
Setembro	0,719
Outubro	0,563
Novembro	0,594
Dezembro	0,656

O Estimador Proposto no Capítulo 5

No Capítulo 5, o previsor obtido através do modelo de Filtro de Kalman para o caso de variâncias desconhecidas é dado por:

$$\tilde{\theta}_2 = q \cdot \hat{\theta}_2 + (1-q) \beta \tilde{\theta}_1, \quad (7.2)$$

onde

$$q = \frac{(v_1+2) \cdot S_2(1)}{(v_1+2) S_2(1) + (v_2+2) S_1(1)}, \quad (7.3)$$

$$S_1(1) = (\hat{\theta}_1 - \tilde{\theta}_1)^2 + \frac{1}{b_1} \quad \text{e} \quad S_2(1) = (\tilde{\theta}_1 - \beta \tilde{\theta}_0)^2 + \frac{1}{b_2} \quad (7.4)$$

O modelo utilizado para este exemplo foi:

$$\hat{\theta}_t = \theta_t + e_t, \quad e_t \sim N(0, B_t \sigma^2)$$

$$\theta_t = \theta_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, V),$$

ou seja, $\beta = 1$.

Se o "chute" inicial for considerado como a média da amostra no mês anterior ao de interesse, $S_1(1)$ se reduz a $\frac{1}{b_1}$ e $S_2(1)$ a $\frac{1}{b_2}$.

É importante salientar que entre dois fechamentos consecutivos, os questionários continuam sendo recebidos e a previsão correspondente ao mês anterior, que será utilizada para o cálculo da previsão do mês de referência, pode ser "melhorada" pela incorporação destes novos questionários.

Vale lembrar que v_1, b_1 e v_2, b_2 são os parâmetros das distribuições a priori de $\frac{1}{B_t \sigma^2}$ e $\frac{1}{V}$, que devem ser escolhidos pelo pesquisador. Para efeito de exercício e aplicação do previsor aos dados, foram escolhidos, arbitrariamente, para q , os seguintes valores: 0,1; 0,2; 0,5; 0,8 e 0,9. Posteriormente, a escolha desses valores e os resultados obtidos, apresentados na tabela 7.2, serão comentados.

Tabela 7.2.-Valores de Θ e suas estimativas para os meses de fevereiro/86 a dezembro/86.

Mês	Θ	Método atual $\hat{Y}_m$	$\hat{\Theta}^*$ (Cap. 4)	$\hat{\Theta}$ (Filtro de Kalman - Cap. 5)				
				$q = 0,1$	$q = 0,2$	$q = 0,5$	$q = 0,8$	$q = 0,9$
Fevereiro	797,22	1.021,25	797,44	817,18	839,85	907,88	975,90	998,58
Março	799,88	1.004,27	798,32	817,93	838,63	900,75	962,86	983,57
Abril	798,09	950,27	797,35	814,92	829,96	875,08	920,19	935,23
Maiο	797,53	886,64	802,42	806,95	815,80	842,37	868,93	877,79
Junho	790,13	966,10	791,24	814,39	831,24	881,82	932,39	949,24
Julho	805,56	873,86	797,08	798,50	806,88	832,00	857,11	865,49
Agosto	792,41	926,95	801,94	817,70	829,84	866,26	902,67	914,81
Setembro	781,91	900,35	787,55	803,20	814,00	846,38	878,76	889,56
Outubro	782,19	897,44	778,26	793,46	805,02	839,68	874,33	885,89
Novembro	788,09	978,00	788,78	801,77	821,35	880,10	938,84	958,42
Dezembro	782,53	947,38	778,78	804,02	819,95	867,74	915,52	931,45

Comentários

1. O método atual sempre superestima a média. Isto é uma particularidade desta pesquisa, pois o questionário envolvido é bastante extenso e as empresas maiores tem mais recursos para repondê-lo com maior rapidez.

2. O estimador $\hat{\theta}^*$ (Capítulo 4) "melhora" o método atual, uma vez que estima o valor correspondente a cada empresa faltante (geralmente as menores), e incorpora essas informações à estimativa de θ .
3. O previsor $\tilde{\theta}$ (Filtro de Kalman - Capítulo 5) também "melhora" o método atual porque utiliza as informações do mês anterior, que, para este caso particular, pouco variam de um mês para outro.
4. O previsor $\tilde{\theta}$ foi calculado para cinco valores diferentes de q . Pode-se observar que quanto menor o valor de q , melhor é a estimativa obtida. Isto já era esperado, uma vez que θ_t varia pouco de um mês para outro e que o painel no instante anterior já está completo.
5. Uma forma natural para a escolha dos parâmetros das priors (quando $v_1 = v_2$) é a seguinte: considere o modelo

$$\begin{aligned}\hat{\theta}_t &= \theta_t + e_t & e_t &\sim N(0, \frac{1}{\sigma_1^2}) \\ \theta_t &= \theta_{t-1} + \varepsilon_t & \varepsilon_t &\sim N(0, \frac{1}{\sigma_2^2}).\end{aligned}$$

Como se tem a informação adicional que θ_t é aproximadamente igual a θ_{t-1} , isto significa que, em geral, $|e_t|$ é maior que $|\varepsilon_t|$, ou seja, a precisão de e_t é menor que a precisão de ε_t . Considerando $v_1 = v_2$, b_1 e b_2 são proporcionais às médias das precisões de e_t e ε_t , respectivamente, e então, b_1 deve ser bem menor que b_2 .

Por outro lado, tem-se que $q = \frac{b_1}{b_1+b_2}$ e, portanto, escolhendo b_1 bem menor do que b_2 , o valor de q será o mais razoável, conforme sugestão dos dados e da informação de que θ_t varia muito pouco em relação a θ_{t-1} .

6. Uma suposição para a utilização do previsor $\tilde{\theta}$ é a de que os dados provêm de uma distribuição Normal. Um teste foi aplicado e seu nível descritivo ficou em torno de 5,2%.

7.3. Um Exemplo com Dados Simulados

No Capítulo 6, o estimador proposto é dado por:

$$\hat{\theta}_t(j) = A\mu + B\hat{\theta}(j) + C \sum_{\substack{i=1 \\ i \neq j}}^t \hat{\theta}(i), \quad (7.5)$$

onde $\hat{\theta}(j)$ é a média amostral no instante j e A , B e C são dados por (6.10), (6.11) e (6.12), respectivamente.

Para a obtenção dessas fórmulas, foi suposto que ρ , n e λ_{ij} (para quaisquer $i, j \in P_2(T)$) são constantes.

Três exemplos serão apresentados, como ilustração, e a fim de termos as suposições acima satisfeitas, as populações foram geradas da seguinte maneira:

- 1) A correlação entre o mesmo elemento, para dois instantes diferentes, foi de $\rho = 0,10$, $\rho = 0,50$ e $\rho = 0,90$, para os exemplos 1, 2 e 3, respectivamente.

2) Foram considerados 11 instantes e em cada instante foi gerada uma população com $N = 550$ elementos, ou seja, $P_2(T) = \{1, 2, \dots, 11\}$ e $P_1(t) = \{1, 2, \dots, 550\}$. Vale salientar que as populações foram geradas segundo uma distribuição Normal, o que não é uma condição para se aplicar esta técnica.

O esquema amostral considerado foi o seguinte: os 550 elementos da população foram divididos em 55 grupos de 10 elementos cada um, de acordo com a ordem em que foram gerados, isto é, os 10 primeiros elementos compõem o grupo 1, os 10 seguintes o grupo 2, e assim por diante. A tabela 7.3 mostra os grupos pertencentes à amostra, em cada instante.

Tabela 7.3. - Grupos pertencentes à amostra para cada instante

INSTANTES										
1	2	3	4	5	6	7	8	9	10	11
1	1	2	3	4	5	6	7	8	9	10
2	11	11	12	13	14	15	16	17	18	19
3	12	20	20	21	22	23	24	25	26	27
4	13	21	28	28	29	30	31	32	33	34
5	14	22	29	35	35	36	37	38	39	40
6	15	23	30	36	41	41	42	43	44	45
7	16	24	31	37	42	46	46	47	48	49
8	17	25	32	38	43	47	50	50	51	52
9	18	26	33	39	44	48	51	53	53	54
10	19	27	34	40	45	49	52	54	55	55

Como cada grupo tem 10 elementos, em cada um dos instantes a amostra tem $n = 100$ elementos. A amostra global cor-

responde a um painel rotativo, onde a proporção de elementos comuns à amostra entre os instantes i e j é constante, ou seja, $\lambda_{ij} = \lambda = 0,10, \forall i, j \in P_2(T)$.

As tabelas 7.4, 7.5 e 7.6 são exemplos de três populações geradas com médias $\mu = 0$, variâncias $10/9$, 2 e 10 e sementes $3,2$; $4,925$ e $1,9$, respectivamente.

Cada uma das tabelas contém, para cada um dos 11 instantes, a média da população, a média amostral e o estimador obtido através do Método de Mínimos Quadrados, quando os parâmetros são variáveis aleatórias.

Tabela 7.4. - Média da População, Média da Amostra e Estimador de Mínimos Quadrados
(Exemplo 1: $\mu = 0$, $\sigma^2 = 10/9$, $\rho = 0,10$, $\lambda = 0,10$ e Semente = $3,2$).

Instante (i)	$\Theta(i)$	$\hat{\Theta}(i)$	$\tilde{\Theta}(i)$
1	-0,078	0,028	0
2	-0,023	-0,058	-0,014
3	0,023	-0,135	-0,027
4	-0,035	-0,022	-0,008
5	-0,007	0,064	0,006
6	-0,056	-0,142	-0,028
7	-0,099	-0,040	-0,011
8	-0,003	-0,009	-0,006
9	-0,036	-0,204	-0,038
10	0,064	0,095	0,011
11	0,076	0,119	0,015

Tabela 7.5. - Média da população, média da amostra e estimador de mínimos quadrados
(Exemplo 2: $\mu = 0$, $\sigma^2 = 2$, $\rho = 0,50$, $\lambda = 0,10$ e Semente = 4,925).

Instante (i)	$\Theta(i)$	$\hat{\Theta}(i)$	$\tilde{\Theta}(i)$
1	-0,042	-0,194	-0,047
2	-0,076	0,021	-0,026
3	0,048	0,152	-0,014
4	-0,044	0,037	-0,025
5	0,012	-0,314	-0,058
6	-0,052	-0,138	-0,041
7	-0,042	-0,038	-0,032
8	-0,054	0,071	-0,021
9	-0,104	0,083	-0,020
10	-0,051	-0,116	-0,039
11	-0,004	-0,053	-0,033

Tabela 7.6. - Média da população, média da amostra e estimador de mínimos quadrados
(Exemplo 3: $\mu = 0$, $\sigma^2 = 10$, $\rho = 0,90$, $\lambda = 0,10$ e Semente=1,9).

Instante (i)	$\Theta(i)$	$\hat{\Theta}(i)$	$\tilde{\Theta}(i)$
1	0,159	0,513	0,155
2	0,154	-0,187	0,141
3	0,133	0,389	0,152
4	0,147	0,066	0,146
5	0,137	0,398	0,152
6	0,111	-0,080	0,143
7	0,096	0,070	0,146
8	0,184	0,159	0,147
9	0,164	0,019	0,145
10	0,166	0,201	0,148
11	0,181	0,278	0,150

Os gráficos 7.1, 7.2 e 7.3 mostram, para os 11 ins -
tantes, o comportamento da média populacional, da média amos -
tral e do estimador de mínimos quadrados apresentados nas tabe -
las 7.4, 7.5 e 7.6, respectivamente.

Comentários

1. Nos três exemplos, em geral, o estimador de mínimos quadrados se encontra mais próximo da média populacional do que a média amostral.
2. Quando o parâmetro populacional apresenta um "pico" em algum instante (por exemplo, no instante 3 do gráfico 7.2), o estimador de mínimos quadrados não consegue detectá-lo, pois seu comportamento é "alisado" por utilizar as informações de todos os instantes para estimar a média daquele particular instante.
3. Para o estimador de mínimos quadrados, foi considerado o caso em que μ é conhecida. Se μ é desconhecida, o estimador se reduz à média amostral.

4. Quanto maior for o valor de ρ , mais próximo da média populacional estará o estimador proposto. Isto é intuitivo pois, como esse estimador utiliza as informações de to dos os instantes, se a correlação é maior, melhor será o estimador.

oOo

A P Ê N D I C E

Teorema I:

O conjunto de observações $y_i(t)$, $i \in S(t)$, $t=1,2,\dots,\tau$ pode ser visualizado como uma matriz incompleta P , com τ linhas e tantas colunas quanto o número de unidades distintas observadas. Se o número de unidades distintas observadas é finito, então P pode ser dividida em um número finito de submatrizes retangulares (cada uma, matriz completa de $y_i(t)$ valores) tal que os pesos associados com os valores de $y_i(t)$ em qualquer uma das linhas de uma dada submatriz são todos iguais na expressão do estimador de variância mínima.

Teorema II:

Seja X o conjunto de observações $x_1, x_2, \dots, x_n$, cuja distribuição depende de um parâmetro desconhecido θ . Seja $\delta(X)$ um estimador não viciado de θ com $\text{Var}(\delta(X)) < \infty$ para todo θ . Seja $\delta_0(X)$ algum estimador não viciado, não trivial, de zero, com variância finita (em toda parte). Então uma condição necessária e suficiente para $\delta(X)$ ter variância mínima é que $\text{Cov}(\delta(X), \delta_0(X)) = 0$. Além disso, se $\delta_1(X)$ é algum outro estimador não viciado de θ , com variância finita (em toda parte) então $\text{Var}(\delta(X)) = \text{Cov}(\delta(X), \delta_1(X))$.

Teorema III: (Gauss-Markov generalizado)

Seja $\underline{Y} = \underline{X}\underline{\beta} + \underline{e}$, $E(\underline{e}) = \underline{0}$ e $V(\underline{e}) = \sigma^2 \underline{W}$, $U(\underline{X})$ o espaço vetorial gerado pelas colunas de $\underline{X}$, $U(\underline{W})$ o espaço vetorial gerado pelas colunas de $\underline{W}$ e $U(\underline{X}) \subset U(\underline{W})$. Tem-se que:

- i) o estimador não viciado e de variância mínima para uma função estimável $\underline{\ell}'\underline{\beta}$ é dado por $\underline{\ell}'\hat{\underline{\beta}}$, onde

$$\hat{\underline{\beta}} = (\underline{X}'\underline{W}^{-1}\underline{X})^{-1}\underline{X}'\underline{W}^{-1}\underline{Y};$$

- ii) um estimador não viciado de σ^2 é dado por

$$\hat{\sigma}^2 = \frac{(\underline{Y} - \underline{X}\hat{\underline{\beta}})' \underline{W}^{-1} (\underline{Y} - \underline{X}\hat{\underline{\beta}})}{n-p}$$

onde n é o posto de $\underline{W}$ e p é o posto de $\underline{X}$.

Teorema IV: (Gauss-Markov para Populações Finitas)

Seja $\theta_L(\underline{Y}) = \underline{\ell}'\underline{Y}$ e $U_{\underline{\ell}}$ a classe de todos os previsores lineares ζ - não viciados para $\theta_L(\underline{Y})$. Entre os previsores pertencentes à classe $U_{\underline{\ell}}$, o previsor linear ζ -ótimo é dado por

$$\hat{\theta}_G(\underline{Y}_S) = \underline{\ell}'\underline{Y}_S + \underline{\ell}'\underline{X}_S \hat{\underline{\beta}} \quad \text{onde} \quad \hat{\underline{\beta}} = (\underline{X}_S' \underline{W}_S^{-1} \underline{X}_S)^{-1} \underline{X}_S' \underline{W}_S^{-1} \underline{Y}_S.$$

Resultado 1:

Quando F é uma função de distribuição de uma variável aleatória discreta X , a integral reduz-se a uma série: se

$$P(X=x_i) = p(x_i) > 0 \quad \text{e} \quad \sum_i p(x_i) = 1,$$

isto é, se p é uma função de probabilidade de X , então $p(x_i)$ é o salto de F em x_i e

$$\int_a^b \phi(x) dF(x) = \sum_i \phi(x_i) p(x_i).$$

Resultado 2:

A integral é linear, tanto no integrando como no integrador. Em outras palavras, para $\phi(x) = \alpha f(x) + \beta g(x)$ tem-se

$$\int_a^b \phi(x) dF(x) = \alpha \int_a^b f(x) dF(x) + \beta \int_a^b g(x) dF(x)$$

e para $H(x) = \alpha F(x) + \beta G(x)$ tem-se

$$\int_a^b \phi(x) dH(x) = \alpha \int_a^b \phi(x) dF(x) + \beta \int_a^b \phi(x) dG(x)$$

Resultado 3:

$$\begin{aligned} \int_0^T c(t) \underset{\sim}{\ell}'_N \underset{\sim}{X}(t) \underset{\sim}{\beta} dF(t) &= \int_0^T c(t) \sum_{k=1}^N \ell_{k \underset{\sim}{i}=1}^p X_{ki}(t) \beta_{\underset{\sim}{i}} dF(t) = \\ &= \sum_{k=1}^N \ell_{k \underset{\sim}{i}=1}^p \int_0^T c(t) X_{ki}(t) dF(t) \beta_{\underset{\sim}{i}} = \\ &= \underset{\sim}{\ell}'_N \left(\int_0^T c(t) \underset{\sim}{X}(t) dF(t) \right) \underset{\sim}{\beta} \quad , \end{aligned}$$

onde

$$\int_0^T c(t) \underset{\sim}{X}(t) dF(t) = \left(\int_0^T c(t) X_{ki}(t) dF(t) \right)_{\substack{1 \leq k \leq N \\ 1 \leq i \leq p}}$$

Resultado 4:

Seja

$$A_{(n \times n)} = \begin{pmatrix} a & b & b & \dots & b \\ b & a & b & \dots & b \\ \vdots & \vdots & \vdots & & \vdots \\ b & b & b & \dots & a \end{pmatrix}$$

então a inversa de A é dada pela matriz cujos elementos são:

$$a_{ii} = \frac{a + (n-2)b}{(a + (n-1)b)(a-b)}$$

e

$$a_{ij} = \frac{-b}{(a + (n-1)b)(a-b)} \quad \text{se } i \neq j$$

Resultado 5:

$$\tilde{Y}(t) \sim N(\mu \tilde{1}_N; \sigma^2 \tilde{I})$$

$$\text{Cov}[Y_k(t), Y_{k'}(t)] = \begin{cases} \sigma^2 \rho & , k = k' , t \neq t' \\ 0 & , k \neq k' , t \neq t' \end{cases}$$

σ^2 e ρ : conhecidos

μ : desconhecido.

Seja $\theta(t) = \frac{1}{N} \tilde{1}'_N \tilde{Y}(t)$ $\theta_t = \sum_{j=1}^t \lambda_j \theta(j)$

$$\hat{\theta}(t) = \frac{1}{n} \sum_{k \in S_t} Y_k(t) \quad \hat{\theta}_t = \sum_{j=1}^t \lambda_j \hat{\theta}(j)$$

$f = \frac{n}{N}$: fração amostral.

λ_{ij} : proporção dos elementos comuns às amostras nos instantes i e j .

i) $E\{\theta(t)\} = E\left\{\frac{1}{N} \tilde{1}'_N \tilde{Y}(t)\right\} = \frac{1}{N} \tilde{1}'_N E\{\tilde{Y}(t)\} = \frac{1}{N} N\mu = \mu$

ii) $\text{Var}\{\theta(t)\} = \text{Var}\left\{\frac{1}{N} \tilde{1}'_N \tilde{Y}(t)\right\} = \frac{1}{N^2} \tilde{1}'_N \sigma^2 \tilde{I} \tilde{1}_N = \frac{\sigma^2}{N}$

iii) $E\{\theta_t\} = E\left\{\sum_{j=1}^t \lambda_j \theta(j)\right\} = \sum_{j=1}^t \lambda_j E\{\theta(j)\} = \mu \sum_{j=1}^t \lambda_j$

$$\begin{aligned}
 \text{iv) } \text{Var}\{\theta_t\} &= \text{Var}\left\{\sum_{j=1}^t \ell_j \theta(j)\right\} = \text{Var}\left\{\sum_{j=1}^t \ell_j \frac{1}{N} \mathbf{1}_N' \mathbf{Y}(j)\right\} = \\
 &= \frac{1}{N^2} \left\{ \sum_{j=1}^t \ell_j^2 \mathbf{1}_N' \sigma^2 \mathbf{I} \mathbf{1}_N + 2 \sum_{j < i} \ell_j \ell_i \text{Cov}(\mathbf{1}_N' \mathbf{Y}(j); \mathbf{1}_N' \mathbf{Y}(i)) \right\} = \\
 &= \frac{1}{N^2} \left\{ \sum_{j=1}^t \ell_j^2 N \sigma^2 + 2 \sum_{j < i} \ell_j \ell_i N \sigma^2 \rho \right\} = \\
 &= \frac{\sigma^2}{N} \left\{ \sum_{j=1}^t \ell_j^2 + 2 \rho \sum_{j < i} \ell_j \ell_i \right\} = \Sigma_{22}
 \end{aligned}$$

$$\text{v) } E\{\hat{\theta}(t)\} = E\left\{\frac{1}{n} \sum_{k \in S_t} Y_k(t)\right\} = \frac{1}{n} \sum_{k \in S_t} E\{Y_k(t)\} = \frac{1}{n} n \mu = \mu$$

$$\begin{aligned}
 \text{vi) } \text{Var}\{\hat{\theta}(t)\} &= \text{Var}\left\{\frac{1}{n} \sum_{k \in S_t} Y_k(t)\right\} = \\
 &= \frac{1}{n^2} \left\{ \sum_{k \in S_t} \text{Var}\{Y_k(t)\} + \sum_{k \neq k'} \text{Cov}(Y_k(t), Y_{k'}(t)) \right\} = \\
 &= \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}
 \end{aligned}$$

$$\text{vii) } E\{\hat{\theta}_t\} = E\left\{\sum_{j=1}^t \ell_j \hat{\theta}(j)\right\} = \sum_{j=1}^t \ell_j E\{\hat{\theta}(j)\} = \mu \sum_{j=1}^t \ell_j$$

$$\begin{aligned}
 \text{viii) } \text{Var}\{\hat{\theta}_t\} &= \text{Var}\left\{\sum_{j=1}^t \ell_j \hat{\theta}(j)\right\} = \\
 &= \sum_{j=1}^t \ell_j^2 \text{Var}\{\hat{\theta}(j)\} + 2 \sum_{j < i} \ell_j \ell_i \text{Cov}\{\hat{\theta}(j), \hat{\theta}(i)\} = \\
 &= \sum_{j=1}^t \ell_j^2 \frac{\sigma^2}{n} + 2 \sum_{j < i} \ell_j \ell_i \text{Cov}\left\{\frac{1}{n} \sum_{k \in S_j} Y_k(j); \frac{1}{n} \sum_{k \in S_i} Y_k(i)\right\} = \\
 &= \sum_{j=1}^t \ell_j^2 \frac{\sigma^2}{n} + 2 \sum_{j < i} \ell_j \ell_i \frac{1}{n^2} m_{ij} \sigma^2 \rho =
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{\sigma^2}{n} \left\{ \sum_{j=1}^t \ell_j^2 + 2\rho \sum_{j<i} \ell_j \ell_i \lambda_{ij} \right\} = \Sigma_{11} \\
 &= \text{Cov}\{\theta_t, \hat{\theta}_t\} = \text{Cov} \left\{ \sum_{j=1}^t \ell_j \frac{1}{N} \sum_{k=1}^N Y_k(j), \sum_{j=1}^t \ell_j \frac{1}{n} \sum_{k \in S_j} Y_k(j) \right\} = \\
 &= \sum_{j=1}^t \ell_j \frac{1}{Nn} \text{nCov}\{Y_k(1), Y_k(j)\} + \sum_{j=1}^t \ell_j \frac{1}{Nn} \text{nCov}\{Y_k(2), Y_k(j)\} + \\
 &\quad \dots + \sum_{j=1}^t \ell_j \frac{1}{Nn} \text{Cov}\{Y_k(t), Y_k(j)\} = \\
 &= \sum_{j=1}^t \ell_j^2 \frac{\sigma^2}{N} + 2 \sum_{j<i} \ell_j \ell_i \sigma^2 \rho = \frac{\sigma^2}{N} \left\{ \sum_{j=1}^t \ell_j^2 + 2\rho \sum_{j<i} \ell_j \ell_i \right\} = \\
 &= \Sigma_{12} = \Sigma_{21} = \Sigma_{22} .
 \end{aligned}$$

Como $\underline{Y}(t)$ tem distribuição normal, $\theta(t)$ e $\hat{\theta}(t)$ também tem distribuição normal e, por consequência, θ_t e $\hat{\theta}_t$ também.

Então, a distribuição conjunta de θ_t e $\hat{\theta}_t$ será Normal Bivariada, com os seguintes parâmetros:

$$\begin{pmatrix} \hat{\theta}_t \\ \theta_t \end{pmatrix} \sim N \left(\begin{pmatrix} \mu \sum_{j=1}^t \ell_j \\ \mu \sum_{j=1}^t \ell_j \end{pmatrix} ; \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

Aplicando-se o resultado 6 do apêndice:

$$E\{\hat{\theta}_t / \theta_t\} = \mu \sum_{j=1}^t \ell_j + 1(\theta_t - \mu \sum_{j=1}^t \ell_j) = \theta_t$$

$$\begin{aligned}\text{Var}\{\hat{\theta}_t | \theta_t\} &= \frac{\sigma^2}{n} \left\{ \sum_{j=1}^t \ell_j^2 + 2\rho \sum_{j < i} \ell_j \ell_i \lambda_{ij} \right\} - 1 \frac{\sigma^2}{N} \left\{ \sum_{j=1}^t \ell_j^2 + 2\rho \sum_{j < i} \ell_j \ell_i \right\} \\ &= \sigma^2 \left\{ \left(\frac{1}{n} - \frac{1}{N} \right) \sum_{j=1}^t \ell_j^2 + \frac{2\rho}{n} \left(\sum_{j < i} \ell_j \ell_i \lambda_{ij} - \sum_{j < i} \ell_j \ell_i \frac{n}{N} \right) \right\} = \\ &= \sigma^2 \left\{ \frac{1}{n} (1-f) \sum_{j=1}^t \ell_j^2 + \frac{2\rho}{n} \sum_{j < i} \ell_j \ell_i (\lambda_{ij} - f) \right\}\end{aligned}$$

Resultado 6:

$$\text{Se } \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

então

$$(X_1/X_2 = x_2) \sim N(\mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2); \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21})$$

Resultado 7:

$$\int_0^\infty (\sigma^2)^{-q} \exp(-\sigma^{-2} Q) d\sigma^2 \propto Q^{-(q-1)}$$

Resultado 8:

Teoria de Mínimos Quadrados quando os parâmetros são variáveis aleatórias.

Considere $\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$, onde $\underline{\beta}$ é variável aleatória e

$$E(\underline{\varepsilon}) = \underline{0} \quad , \quad V(\underline{\varepsilon}) = V[\underline{Y}|\underline{\beta}] = \underline{\Sigma}$$

$$E(\underline{\beta}) = \underline{\mu} \quad , \quad V(\underline{\beta}) = \underline{T}$$

O melhor estimador linear da variável aleatória $\underline{P}'\underline{\beta}$ é dado por $\underline{a}_* + \underline{L}_*'\underline{Y}$, onde $\underline{a}_*$ e $\underline{L}_*$ são dados abaixo, e

$$\begin{cases} E(\underline{P}'\underline{\beta} - \underline{a}_* - \underline{L}_*'\underline{Y}) = 0 \\ \text{Var}(\underline{P}'\underline{\beta} - \underline{a}_* - \underline{L}_*'\underline{Y}) \quad \text{é mínima} \end{cases}$$

Caso 1: $\underline{\mu}$ é conhecida

$$\underline{L}_* = \underline{\Sigma}^{-1}\underline{X}(\underline{T}^{-1} + \underline{X}'\underline{\Sigma}^{-1}\underline{X})^{-1}\underline{P}$$

$$\underline{a}_* = \underline{\mu}'\underline{P} - \underline{\mu}'\underline{X}'\underline{L}_*$$

$$\text{Var}(\underline{P}'\underline{\beta} - \underline{a}_* - \underline{L}_*'\underline{Y}) = \underline{P}'(\underline{T}^{-1} + \underline{X}'\underline{\Sigma}^{-1}\underline{X})^{-1}\underline{P}$$

Caso 2. $\underline{\mu}$ é desconhecida

$$\underline{L}_* = \underline{\Sigma}^{-1}\underline{X}(\underline{X}'\underline{\Sigma}^{-1}\underline{X})^{-1}\underline{P}$$

$$\underline{a}_* = 0$$

$$\text{Var}(\underline{P}'\underline{\beta} - \underline{a}_* - \underline{L}_*'\underline{Y}) = \underline{P}'(\underline{X}'\underline{\Sigma}^{-1}\underline{X})^{-1}\underline{P}$$

Resultado 9

$$\hat{\theta}_t = \mu'_{\tilde{z}} \ell - \mu'_{\tilde{z}} \Sigma^{-1} (T^{-1} + \Sigma^{-1})^{-1} \tilde{\ell} + \tilde{\ell}' (T^{-1} + \Sigma^{-1})^{-1} \Sigma^{-1} \hat{\theta}(t)$$

$$T_{\tilde{z}} = \frac{\sigma^2}{Nn} \begin{pmatrix} n & n\rho & \dots & n\rho \\ n\rho & n & \dots & n\rho \\ \vdots & \vdots & & \vdots \\ n\rho & n\rho & & n \end{pmatrix} \quad \tilde{\ell} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_t \end{pmatrix}$$

$$\Sigma_{\tilde{z}} = \frac{\sigma^2}{Nn} \begin{pmatrix} N-n & (N\lambda-n)\rho & \dots & (N\lambda-n)\rho \\ (N\lambda-n)\rho & N-n & \dots & (N\lambda-n)\rho \\ \vdots & \vdots & & \vdots \\ (N\lambda-n)\rho & (N\lambda-n)\rho & \dots & N-n \end{pmatrix}$$

$$T_{\tilde{z}}^{-1} = \frac{Nn}{\sigma^2} \begin{pmatrix} a & b & \dots & b \\ b & a & \dots & b \\ \vdots & \vdots & & \vdots \\ b & b & \dots & a \end{pmatrix}$$

$$a = \frac{n+(t-2)n\rho}{(n+(t-1)n\rho)(n-n\rho)} = \frac{1+(t-2)\rho}{n(1+(t-1)\rho)(1-\rho)}$$

$$b = \frac{-\rho}{n(1+(t-1)\rho)(1-\rho)}$$

$$\Sigma_{\tilde{z}}^{-1} = \frac{Nn}{\sigma^2} \begin{pmatrix} c & d & \dots & d \\ d & c & \dots & d \\ \vdots & \vdots & & \vdots \\ d & d & \dots & c \end{pmatrix}$$

$$c = \frac{N-n+(t-2) (N\lambda-n) \rho}{(N-n+(t-1) (N\lambda-n) \rho) (N-n-(N\lambda-n) \rho)}$$

$$d = \frac{-(N\lambda-n) \rho}{(N-n+(t-1) (N\lambda-n) \rho) (N-n-(N\lambda-n) \rho)}$$

$$(\tilde{T}^{-1} + \tilde{\Sigma}^{-1}) = \frac{Nn}{\sigma^2} \begin{pmatrix} a+c & b+d \dots b+d \\ b+d & a+c \dots b+d \\ \vdots & \vdots & \vdots \\ b+d & b+d \dots a+c \end{pmatrix}$$

$$(\tilde{T}^{-1} + \tilde{\Sigma}^{-1}) = \frac{\sigma^2}{Nn} \begin{pmatrix} h & g \dots g \\ g & h \dots g \\ \vdots & \vdots & \vdots \\ g & g \dots h \end{pmatrix}$$

$$h = \frac{a+c + (t-2) (b+d)}{(a+c + (t-1) (b+d)) (a+c-(b+d))}$$

$$g = \frac{-(b+d)}{(a+c + (t-1) (b+d)) (a+c-(b+d))}$$

$$(\tilde{T}^{-1} + \tilde{\Sigma}^{-1})^{-1} \tilde{\Sigma}^{-1} = [a_{ij}]$$

$$a_{ij} = \begin{cases} hc + (t-1)gd & , \quad i = j \\ hd + gc + (t-2)gd & , \quad i \neq j \end{cases}$$

$$\mu'_{\sim} \ell_{\sim} = \mu \sum_{j=1}^t a_j$$

$$\mu'_{\sim} \Sigma^{-1} (T^{-1} + \Sigma^{-1})^{-1} \ell_{\sim} = \mu [hc + (t-1)hd + (t-1)gc + (t-1)^2 gd] \sum_{j=1}^t a_j$$

$$\begin{aligned} \ell'_{\sim} (T^{-1} + \Sigma^{-1})^{-1} \Sigma^{-1} \hat{\theta}(t) &= (hc + (t-1)gd) \sum_{j=1}^t a_j \hat{\theta}(j) + \\ &+ (hd + gc + (t-2)gd) \sum_{j=1}^t \hat{\theta}(j) \sum_{\substack{i=1 \\ i \neq j}}^t a_i \end{aligned}$$

mas

I)

$$\begin{aligned} hc + (t-1)hd + (t-1)gc + (t-1)^2 gd &= \\ &= h(c + (t-1)d) + g(t-1)(c + (t-1)d) = \\ &= (h + (t-1)g)(c + (t-1)d) = \\ &= \frac{a+c+(t-2)(b+d) - (t-1)(b+d)}{(a+c+(t-1)(b+d))(a+c-(b+d))} (c + (t-1)d) = \\ &= \frac{c + (t-1)d}{a+c+(t-1)(b+d)} = (*) \end{aligned}$$

mas

$$a + (t-1)b = \frac{1 + (t-2)\rho - (t-1)\rho}{n(1 + (t-1)\rho)(1-\rho)} = \frac{1}{n(1 + (t-1)\rho)}$$

e

$$c + (t-1)d = \frac{1}{(N-n + (t-1)(N\lambda - n)\rho)}$$

então,

$$\begin{aligned}
 (*) &= \frac{\frac{1}{N-n + (t-1) (N\lambda-n)\rho}}{\frac{1}{n(1+(t-1)\rho)} + \frac{1}{N-n+(t-1) (N\lambda-n)\rho}} = \\
 &= \frac{n(1+(t-1)\rho)}{n(1+(t-1)\rho) + N-n + (t-1) (N\lambda-n)\rho} = \\
 &= \frac{n(1+(t-1)\rho)}{n+n(t-1)\rho+N-n+(t-1)N\lambda\rho-(t-1)n\rho} = \\
 &= \frac{n(1+(t-1)\rho)}{N(1+(t-1)\lambda\rho)} = 1-A
 \end{aligned}$$

II)

$$\begin{aligned}
 hc+(t-1)gd &= \frac{(a+c+(t-2)(b+d)c-(t-1)(b+d)d}{(a+c+(t-1)(b+d))(a+c-(b+d))} = \\
 &= \frac{(a+(t-2)b+c+(t-2)d)c-(t-1)(b+d)d}{(a+(t-1)b+c+(t-1)d)(a-b+c-d)} = (**)
 \end{aligned}$$

mas

$$a+(t-2)b = \frac{1}{n(1+(t-1)\rho)(1-\rho)} ,$$

$$c + (t-2)d = \frac{N-n}{(N-n+(t-1)(N\lambda-n)\rho)(N-n-(N\lambda-n)\rho)} ,$$

$$a - b = \frac{1}{n(1-\rho)} \quad c - d = \frac{1}{N-n-(N\lambda-n)\rho}$$

e então

$$\begin{aligned}
 (**) &= \frac{\frac{c+(t-1)\rho d}{n(1+(t-1)\rho)(1-\rho)} + \frac{(N-n)c+(t-1)(N\lambda-n)\rho d}{(N-n+(t-1)(N\lambda-n)\rho)(N-n-(N\lambda-n)\rho)}}{\frac{N+(t-1)N\lambda\rho}{n(1+(t-1)\rho)(N-n+(t-1)(N\lambda-n)\rho)} \frac{N-N\lambda\rho}{n(1-\rho)(N-n-(N\lambda-n)\rho)}} = \\
 &= \frac{(c+(t-1)\rho d)(N-n+(t-1)(N\lambda-n)\rho)(N-n-(N\lambda-n)\rho)n}{(N+(t-1)N\lambda\rho)(N-N\lambda\rho)} + \\
 &+ \frac{n(1+(t-1)\rho)(1-\rho)[(N-n)c+(t-1)(N\lambda-n)\rho d]}{(N+(t-1)N\lambda\rho)(N-N\lambda\rho)} n = \\
 &= n \frac{[(N-n+(t-1)(N\lambda-n)\rho)(N-n-(N\lambda-n)\rho) + (N-n)n(1+(t-1)\rho)(1-\rho)](N-n+(t-2)(N\lambda-n)\rho)}{(N+(t-1)N\lambda\rho)(N-N\lambda\rho)(N-n+(t-1)(N\lambda-n)\rho)(N-n-(N\lambda-n)\rho)} + \\
 &- n \frac{[(N-n+(t-1)(N\lambda-n)\rho)(N-n-(N\lambda-n)\rho)(t-1)\rho + n(1+(t-1)\rho)(1-\rho)(t-1)(N\lambda-n)\rho](N\lambda-n)\rho}{(N+(t-1)N\lambda\rho)(N-N\lambda\rho)(N-n+(t-1)(N\lambda-n)\rho)(N-n-(N\lambda-n)\rho)} = \\
 &= n \frac{(N-n+(t-2)(N\lambda-n)\rho) - (t-1)(N\lambda-n)\rho^2}{N^2(1+(t-1)\lambda\rho)(1-\lambda\rho)} + \\
 &+ n \frac{(N-n)n(1+(t-1)\rho)(1-\rho)(N-n+(t-2)(N\lambda-n)\rho - n(1+(t-1)\rho)(1-\rho)(t-1)(N\lambda-n)\rho^2)}{N^2(1+(t-1)\lambda\rho)(1-\lambda\rho)(N-n+(t-1)(N\lambda-n)\rho)(N-n-(N\lambda-n)\rho)} = \\
 &= n \frac{N-n+(N\lambda-n)((t-2)\rho - (t-1)\rho^2)}{N^2(1+(t-1)\lambda\rho)(1-\lambda\rho)} + \\
 &+ n \frac{\frac{n(1+(t-1)\rho)(1-\rho)[(N-n+(t-2)(N\lambda-n)\rho)(N-n) - (t-1)(N\lambda-n)\rho^2]}{N^2(1+(t-1)\lambda\rho)(1-\lambda\rho)[(N-n) + (t-2)(N\lambda-n)\rho(N-n) - (t-1)(N\lambda-n)\rho^2]}}{N^2(1+(t-1)\lambda\rho)(1-\lambda\rho)} = \\
 &= n \frac{N-n+(N\lambda-n)((t-2)\rho - (t-1)\rho^2) + n(1+(t-1)\rho)(1-\rho)}{N^2(1+(t-1)\lambda\rho)(1-\lambda\rho)} = \\
 &= n \frac{N-n+(N\lambda-n)(t-2)\rho - (N\lambda-n)(t-1)\rho^2 + n-n\rho + n(t-1)\rho - n(t-1)\rho^2}{N^2(1+(t-1)\lambda\rho)(1-\lambda\rho)} = \\
 &= n \frac{N+N\lambda(t-2)\rho - n(t-2)\rho - N\lambda(t-1)\rho^2 + n(t-1)\rho^2 + (t-2)n\rho - (t-1)n\rho^2}{N^2(1+(t-1)\lambda\rho)(1-\lambda\rho)} = \\
 &= \frac{n}{N} \frac{1+(t-2)\lambda\rho - (t-1)\lambda\rho^2}{(1+(t-1)\lambda\rho)(1-\lambda\rho)} = B
 \end{aligned}$$

III)

$$\begin{aligned}
 hd+gc+(t-2)gd &= \frac{(a+c)d-(b+d)c}{(a+c+(t-1)(b+d))(a+c-(b+d))} = \\
 &= \frac{ad+cd-bc-dc}{(a+(t-1)b+c+(t-1)d)(a-b+c-d)} = \\
 &= \frac{- (1+(t-2)\rho)(N\lambda-n)\rho + \rho(N-n+(t-2)(N\lambda-n)\rho)}{n(1+(t-1)\rho)(1-\rho)(N-n+(t-1)(N\lambda-n)\rho)(N-n-(N\lambda-n)\rho)} = \\
 &= \frac{(N-n+(t-1)(N\lambda-n)\rho)+n(1+(t-1)\rho)}{n(1+(t-1)\rho)(N-n+(t-1)(N\lambda-n)\rho)} \cdot \frac{n(1-\rho)+N-n-(N\lambda-n)\rho}{n(1-\rho)(N-n-(N\lambda-n)\rho)} = \\
 &= -n \frac{(1+(t-2)\rho)(N\lambda-n)\rho - (N-n+(t-2)(N\lambda-n)\rho)\rho}{(N+(t-1)N\lambda\rho)(N-N\lambda\rho)} = \\
 &= - \frac{n}{N} \rho \frac{N\lambda-n+(t-2)(N\lambda-n)\rho - N+n-(t-2)(N\lambda-n)\rho}{N(1+(t-1)\lambda\rho)(1-\lambda\rho)} = \\
 &= \frac{n}{N} \rho \frac{1-\lambda}{(1+(t-1)\lambda\rho)(1-\lambda\rho)} = C
 \end{aligned}$$

Resultado 10:

$$\hat{\theta}^* = a \cdot \bar{Y}_m(0) + b \bar{Y}_u(0) + c \bar{Y}_m(1)$$

Como $E_{\zeta}[\hat{\theta}^*] = E_{\zeta}[\theta(1)]$, então

$$a E_{\zeta}[\theta(0)] + b E_{\zeta}[\theta(0)] + c E_{\zeta}[\theta(1)] = E_{\zeta}[\theta(1)]$$

o que implica em $c=1$ e $a=-b$. Logo $\hat{\theta}^*$ pode ser escrito como

$$a \bar{Y}_m(0) - a \bar{Y}_u(0) + \bar{Y}_m(1).$$

Usando o Teorema II do apêndice, tem-se

$$\text{Cov}_{\zeta}[\hat{\theta}^*, \hat{\theta}_0] = 0 , \quad \text{ou,}$$

$$\text{Cov}_{\zeta}[a \bar{Y}_m(0) - a \bar{Y}_u(0) + \bar{Y}_m(1) , \bar{Y}_m(0) - \bar{Y}_u(0)] =$$

$$= a \frac{1}{m} \sigma_0^2 + a \frac{1}{u} \sigma_0^2 + \frac{1}{m} \rho \sigma_0 \sigma_1 =$$

$$= a \sigma_0^2 \frac{u+m}{mu} + \frac{1}{m} \rho \sigma_0 \sigma_1 = 0 .$$

Logo,

$$a = -\rho \frac{\sigma_0 \sigma_1}{m} \frac{mu}{n \sigma_0^2} = -\rho \frac{\sigma_1}{\sigma_0} \frac{u}{n} = -\rho K(1-\lambda) ,$$

ou seja,

$$\hat{\theta}^* = \rho K(1-\lambda) (\bar{Y}_u(0) - \bar{Y}_m(0)) + \bar{Y}_m(1) .$$

Ainda pelo Teorema II (apêndice), pode-se calcular a variância de $\hat{\theta}^*$.

Considerando-se $\bar{Y}_m(1)$, que é outro estimador não viado para $E_{\zeta}[\theta]$, $\text{Var}_{\zeta}[\hat{\theta}^*] = \text{Cov}_{\zeta}[\hat{\theta}^*, \bar{Y}_m(1)]$, ou seja,

$$\begin{aligned}\text{Var}_{\zeta}[\hat{\theta}^*] &= \text{Cov}_{\zeta}[\rho K(1-\lambda)(\bar{Y}_u(0) - \bar{Y}_m(0)) + \bar{Y}_m(1), \bar{Y}_m(1)] = \\ &= -\rho K(1-\lambda) \frac{1}{m} \rho \sigma_0 \sigma_1 + \frac{1}{m} \sigma_1^2 = \frac{-\rho^2 \sigma_1^2 (1-\lambda)}{m} + \frac{\sigma_1^2}{m}\end{aligned}$$

ou

$$\text{Var}_{\zeta}[\hat{\theta}^*] = \frac{\sigma_1^2}{m} [1 - \rho^2 (1-\lambda)]$$

Resultado 11

$$\hat{\theta}^* = a\bar{Y}_m(0) + b\bar{Y}_u(0) + c\bar{Y}_m(1) + d\bar{Y}_u(1)$$

$$E_{\zeta}[\hat{\theta}^*] = aE_{\zeta}[\theta(0)] + bE_{\zeta}[\theta(0)] + cE_{\zeta}[\theta(1)] + dE_{\zeta}[\theta(1)] = E_{\zeta}[\theta(1)]$$

o que implica em $a+b = 0$ e $c+d = 1$, ou seja,

$$\hat{\theta}^* = a\bar{Y}_m(0) - a\bar{Y}_u(0) + c\bar{Y}_m(1) + (1-c)\bar{Y}_u(1)$$

Pelo Teorema II do apêndice,

$$\begin{aligned}\text{Cov}_{\zeta}[\hat{\theta}^*, \hat{\theta}_0] &= \text{Cov}_{\zeta}[a\bar{Y}_m(0) - a\bar{Y}_u(0) + c\bar{Y}_m(1) + (1-c)\bar{Y}_u(1), \bar{Y}_m(0) - \bar{Y}_u(0)] = \\ &= a\frac{1}{m}\sigma_0^2 + a\frac{1}{u}\sigma_0^2 + c\frac{1}{m}\rho\sigma_0\sigma_1 = 0\end{aligned}$$

$$a u \sigma_0^2 + a m \sigma_0^2 + c \rho \sigma_0 \sigma_1 u = 0$$

$$a \sigma_0^2 (u+m) + c \rho \sigma_0 \sigma_1 u = 0 \quad (1)$$

e

$$\begin{aligned}\text{Cov}_{\zeta}[a\bar{Y}_m(0) - a\bar{Y}_u(0) + c\bar{Y}_m(1) + (1-c)\bar{Y}_u(1), \bar{Y}_m(1) - \bar{Y}_u(1)] &= \\ = a\frac{1}{m}\rho\sigma_0\sigma_1 + c\frac{1}{m}\sigma_1^2 - (1-c)\frac{1}{u}\sigma_1^2 &= 0\end{aligned}$$

$$a \rho \sigma_0 \sigma_1 u + c \sigma_1^2 u - (1-c) \sigma_1^2 m = 0$$

$$a \rho \sigma_0 \sigma_1 u + c \sigma_1^2 (u+m) = \sigma_1^2 m \quad (2)$$

Resolvendo o sistema:

$$a = \frac{-c\rho\sigma_0^2\sigma_1^2u}{\sigma_0^2n} = -c\rho K(1-\lambda)$$

$$-c\rho^2\sigma_1^2(1-\lambda)u + c\sigma_1^2n = \sigma_1^2m$$

$$c(n-\rho^2(1-\lambda)u) = m$$

$$c = \frac{m}{n-\rho^2(1-\lambda)u} = \frac{\lambda}{1-(1-\lambda)^2\rho^2} = A$$

$$a = \frac{-\lambda\rho K(1-\lambda)}{1-(1-\lambda)^2\rho^2} = C$$

Considerando $\bar{Y}_u(1)$, que é outro estimador não viciado de $E_\zeta[\theta]$ e aplicando o Teorema II (apêndice)

$$\begin{aligned} \text{Var}_\zeta[\tilde{\theta}^*] &= \text{Cov}_\zeta[\hat{\theta}^*, \bar{Y}_u(1)] = \\ &= \text{Cov}_\zeta[a\bar{Y}_m(0) - a\bar{Y}_u(0) + c\bar{Y}_m(1) + (1-c)\bar{Y}_u(1), \bar{Y}_u(1)] = \\ &= (1-c)\frac{1}{u}\sigma_1^2 = \frac{1-(1-\lambda)^2\rho^2-\lambda}{1-(1-\lambda)^2\rho^2} \frac{\sigma_1^2}{u} = \\ &= \frac{\sigma_1^2(1-\lambda)[1-(1-\lambda)\rho^2]}{u[1-(1-\lambda)^2\rho^2]} = \frac{\sigma_1^2}{n} \left(\frac{1-(1-\lambda)\rho^2}{1-(1-\lambda)^2\rho^2} \right) \end{aligned}$$

oOo

REFERÊNCIAS BIBLIOGRÁFICAS

- BARTLETT, R.F. (1986) Estimating the total of a continuous population. *J. Statist. Planning and Inference*, 13: 51-66.
- BOLFARINE, H. & RODRIGUES, J. (1984) *Teoria da previsão em populações finitas*. Rio de Janeiro, UFRJ - 80 p.
- BOLFARINE, H. & RODRIGUES, J. (1987) A Kalman Filter model for single and two-stage repeated surveys. *Statistics & Probability Letters*, 5: 299-303.
- BOLFARINE, H. & RODRIGUES, J. (1987) A simple projection predictor in finite populations. *Estatística*.
- BROEMELING, L.; DIAZ, J.; YUSOFF, A.M. (1985) Some Bayesian solutions for problems of adaptive estimation in linear dynamic systems. *Comm. Statist. - Theor. Meth.*, 14 (2): 401-418.
- CASSEL, C.M.; SÄRNDAL, C.E.; WRETMAN, J.H. (1977) *Foundations of inference in survey sampling*. New York, John Wiley. 192 p.
- COCHRAN, W.G. (1963) *Sampling techniques*. New York, John Wiley. 330 p.
- DUNCAN, D.B. & HORN, S.D. (1972) Linear dynamic recursive estimation from the viewpoint of regression analysis. *J. Amer. Statist. Assoc.*, 67 (340): 815-821.

- DUNCAN, G.J. & KALTON, G. (1987) Issues of design and analysis of surveys across time. *Internacional Statistical Review*, 55 (1): 97-117.
- ELIAN, S.N. (1984) *Previsão linear em populações finitas*. São Paulo. 172 p. Dissertação (Mestrado) - IME-USP.
- FIROUZABADI, M.D. (1975) *Sampling and time series analysis*. Berkeley. 82 p. Ph.D. Dissertation - University of California.
- FORTNEY, W.G. & MILLER, R.B. (1987) Bayesian analysis in random coefficient m-group regression. *Statistics & Probability Letters*, 5: 135-142.
- GRAHAM, J.E. & RAO, J.N.K. (1964) Rotation designs for sampling on repeated occasions. *J. Amer. Statist. Assoc.*, 59 (306): 492-509.
- HERSON, J. & ROYALL, R.M. (1973) Robust estimation in finite populations I. *J. Amer. Statist. Assoc.*, 68 (344): 880-889.
- JOHNSON, N.L. & KOTZ, S. (1972) *Distributions in statistics: continuous multivariate distributions*. New York, John Wiley. 333 p.
- JONES, R.G. (1980) Best linear unbiased estimators for repeated surveys. *J. Royal Statist. Soc. Ser. B*, 42 (2): 221-226.

- KALMAN, R.E. (1960). A new approach to linear filtering and prediction problems. *Transactions ASME J. Basic Engineering*, 82: 35-45.
- LINDLEY, D.V. & SMITH, A.F.M. (1972) Bayes estimates for the linear model: with discussion. *J. Royal Statist. Soc. Ser. B*, 34 (1): 1-41.
- MEINHOLD, R.J. & SINGPURWALLA, N.D. (1983) Understanding the Kalman Filter. *The American Statistician*, 37 (2): 123-127.
- MORETTIN, P.A. & TOLOI, C.M.C. (1981) *Modelos para previsões de séries temporais*. Poços de Caldas, IMPA - 624 p.
- PATTERSON, H.D. (1950) Sampling on successive occasions with partial replacement of units. *J. Royal Statist. Soc. Ser. B*, 12: 241-255.
- PFEFFERMANN, D. (1984) On extensions of the Gauss-Markov theorem to the case of stochastic regression coefficients. *J. Royal Statist. Soc. Ser. B*, 46 (1): 139-148.
- PRESS, J.S. (1972) *Applied multivariate analysis*. New York, Holt Rinehart and Winston, 521 p.
- RAO, C.R. (1973) *Linear statistical inference and its applications*. New York, John Wiley. 625 p.

SÄRNDAL, C.E. & WRIGHT, R. (1984) Cosmetic forms of estimators in survey sampling. *Scand. J. Statistics*, 11 (3): 146-156.

TAM, S.M. (1986) Optimal prediction in stochastic regression models with application to the analysis of repeated surveys. *Austral. J. Statist.*, 28 (3): 345-353.

oOo