

Presidente da República
Luíz Inácio Lula da Silva

Ministro do Planejamento, Orçamento e Gestão
Paulo Bernardo Silva

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE

Presidente
Eduardo Pereira Nunes

Diretor Executivo
Sérgio da Costa Côrtes

ÓRGÃOS ESPECÍFICOS SINGULARES

Diretoria de Pesquisas
Wasmália Socorro Barata Bivar

Diretoria de Geociências
Guido Gelli

Diretoria de Informática
Luiz Fernando Pinto Mariano

Centro de Documentação e Disseminação de Informações
David Wu Tai

Escola Nacional de Ciências Estatísticas
Pedro Luis do Nascimento Silva

Ministério do Planejamento, Orçamento e Gestão
Instituto Brasileiro de Geografia e Estatística - IBGE

REVISTA BRASILEIRA DE ESTATÍSTICA

volume 66 número 225 janeiro/dezembro 2005

ISSN 0034-7175

R. bras. Estat., Rio de Janeiro, v. 66, n. 225, p. 1-111, jan./dez. 2005

Instituto Brasileiro de Geografia e Estatística - IBGE
Av. Franklin Roosevelt, 166 - Centro - 20021-120 - Rio de Janeiro - RJ - Brasil

© IBGE. 2006

Revista Brasileira de Estatística, ISSN 0034-7175

Órgão oficial do IBGE e da Associação Brasileira de Estatística – ABE.

Publicação semestral que se destina a promover e ampliar o uso de métodos estatísticos (quantitativos) na área das ciências econômicas e sociais, através de divulgação de artigos inéditos.

Temas abordando aspectos do desenvolvimento metodológico serão aceitos, desde que relevantes para os órgãos produtores de estatísticas.

Os originais para publicação deverão ser submetidos em três vias (que não serão devolvidas) para:

Francisco Louzada-Neto

Editor responsável – RBEs – IBGE.

Av. República do Chile, 500 – Centro

20031-170 – Rio de Janeiro, RJ.

Os artigos submetidos às RBEs não devem ter sido publicados ou estar sendo considerados para publicação em outros periódicos.

A Revista não se responsabiliza pelos conceitos emitidos em matéria assinada.

Editor Responsável

Francisco Louzada-Neto (UFSCAR)

Editor de Estatísticas Oficiais

Denise Britz do Nascimento Silva (GAB/IBGE)

Editor de Metodologia

Enrico Antonio Colosimo (UFMG)

Editores Associados

Gilberto Alvarenga Paula (USP)

Dalton Francisco de Andrade (UFSC)

Ismenia Blavatsky de Magalhães (DPE/IBGE)

Helio dos Santos Migon (UFRJ)

Francisco Cribari-Neto (UFPE)

Editoração

Helem Ortega da Silva - Coordenação de Métodos e Qualidade - DPE/COMEQ/IBGE

Impressão

Gráfica Digital/Centro de Documentação e Disseminação de Informações - CDDI/IBGE, em 2004.

Capa

Renato J. Aguiar – Coordenação de *Marketing*/CDDI/IBGE

Ilustração da Capa

Marcos Balster – Coordenação de *Marketing*/CDDI/IBGE

Revista brasileira de estatística/IBGE, - v.1, n.1 (jan./mar.1940), - Rio de Janeiro:IBGE, 1940-

v.

Trimestral (1940-1986), semestral (1987-).

Continuação de: Revista de economia e estatística.

Índices acumulados de autor e assunto publicados no v.43 (1940-1979) e v. 50 (1980-1989).

Co-edição com a Associação Brasileira de Estatística a partir do v.58.

ISSN 0034-7175 = Revista brasileira de estatística.

I. Estatística – Periódicos. I. IBGE. II. Associação Brasileira de Estatística.

IBGE. CDDI. Div. de Biblioteca e Acervos Especiais

CDU 31 (05)

RJ-IBGE/88-05 (rev.98)

PERIÓDICO

Impresso no Brasil/Printed in Brazil

Sumário

Nota do Editor.....	5
Artigos	
Conceitos estatísticos: reflexões.....	7
	<i>Carlos Alberto de Bragança Pereira</i>
Comentário a respeito de "Conceitos Estatísticos: Reflexões", de Carlos Alberto de Bragança Pereira.....	31
	<i>Marco Antônio Rosa Ferreira</i>
Discussão do artigo "Conceitos Estatísticos: Reflexões", de Carlos Alberto de Bragança Pereira.....	36
	<i>Rosângela Helena Loschi</i>
Comentário sobre o artigo "Conceitos Estatísticos: Reflexões", de Carlos Alberto de Bragança Pereira.....	39
	<i>Gilberto A. Paula</i>
Réplica: Carlos Alberto de Bragança Pereira.....	42
Réplica: Marco Antonio Rosa Ferreira.....	45
Tréplica: Carlos Alberto de Bragança Pereira.....	47
Identificação dos fatores determinantes do crescimento em crianças com baixa estatura: uma aplicação do Modelo de Cox estratificado combinado com um modelo para dados longitudinais.....	51
	<i>Magda Carvalho Pires</i> <i>Enrico Antônio Colosimo</i>
Bootstrap estimation of standard errors in seasonal analysis.....	63
	<i>R.P. Mentz</i> <i>N.M. Järma</i> <i>C.I. Martínez</i>
Política editorial.....	109

Nota do Editor

Cada vez mais percebo que a comunidade acadêmica está presenciando uma Era de enorme quantidade de responsabilidades individuais que, direta ou indiretamente, acabam se traduzindo em grandes dificuldades para o cumprimento de prazos.

Do início da gestão do atual corpo editorial da RBEs, perseguimos com afinco o ideal de colocar o periódico em dia, e é com satisfação que comunico aos leitores da RBEs que nosso objetivo está sendo alcançado com este volume. Tenho certeza que com isto estamos incentivando a submissão e colocando o periódico em um outro patamar da divulgação da ciência estatística nacional.

Este volume, também, é especial por incluir um artigo convidado com discussões, o que como havia mencionado na Nota do Editor do volume 224 proporciona aos leitores diferentes visões sobre o mesmo tema.

O artigo convidado é de autoria do Professor Carlos Alberto de Bragança Pereira, um dos ícones da ciência estatística nacional. Professor do IME-USP, com mais de 100 artigos publicados em periódicos, sete livros de sua autoria e quase 30 orientandos de Pós-Graduação (mestrado e doutorado), o Prof. Carlos A. de Bragança Pereira estará completando 60 anos em 2006, dos quais quase dois terços de dedicação à nossa profissão. Reconhecido internacionalmente pelas suas contribuições à ciência estatística, o Prof. Carlos A. de Bragança Pereira nos agracia neste artigo convidado com suas percepções sobre conceitos estatísticos que usualmente são apresentados em sala de aula que podem produzir conclusões incoerentes. Vários exemplos são discutidos com grande didatismo.

O artigo é discutido por três importantes pesquisadores de nossa área. São eles, os professores Gilberto Alvarenga Paula (IME-USP), Marco A. R. Ferreira (UFRJ) e Rosangela Helena Loschi (UFMG).

Também compõe o volume mais dois artigos interessantes. O primeiro de autoria dos pesquisadores Magda Carvalho Pires e Enrico Antonio Colosimo versa sobre a identificação dos fatores determinantes do crescimento em crianças com baixa estatura via modelo de Cox estratificado combinado com um modelo para dados longitudinais. No segundo artigo os pesquisadores R. P. Mentz, N. M. Jarma e C. I. Martinez argumentam que procedimentos *bootstrap* são adequados como uma metodologia alternativa para estimação dos desvios padrão em um particular tipo de séries temporais.

Finalizo minha nota agradecendo imensamente à equipe do IBGE, a nossa secretária Helem Ortega da Silva, e aos membros do corpo editorial da RBEs, particularmente às editoras Ismenia Blavatsky de Magalhães e Denise Britz do Nascimento Silva, pela grande contribuição para que pudéssemos colocar o periódico em dia.

Também agradeço aos pesquisadores Carlos Alberto de Bragança Pereira, Gilberto Alvarenga Paula, Marco A. R. Ferreira e Rosangela Helena Loschi que em meio a enorme quantidade de afazeres se propuseram a escrever e discutir o artigo convidado.

Uma ótima leitura.

Francisco Louzada-Neto
Editor Responsável

Conceitos estatísticos: reflexões

*Carlos Alberto de Bragança Pereira**

Resumo

Este é um artigo sobre conceitos estatísticos que usualmente são apresentados em sala de aula. Vários exemplos são discutidos com o intuito de mostrar que esses conceitos podem produzir conclusões incoerentes. Os exemplos ilustram os seguintes fatos: variáveis aleatórias *iid* são de fato fortemente associadas; probabilidades condicionais podem depender da forma como os condicionantes foram introduzidos; intervalos de confiança podem indicar diminuição de precisão quando a informação é aumentada; testes de significância podem não rejeitar hipóteses impossíveis.

Palavras-chave: Confiança, credibilidade, associação, significância, independência estatística.

1. Introdução

Atualmente, nota-se que o aprendizado de disciplinas básicas, como matemática ou física, tornou-se uma barreira para o aumento desejado de profissionais com formação direcionada aos desafios científicos e tecnológicos do mundo moderno. O desinteresse dos alunos pelas disciplinas técnicas, que exigem um maior poder de raciocínio lógico e abstrato, é observado em vários cantos do mundo. Por outro lado, os valores da sociedade moderna vêm se transformando rapidamente, acompanhando a velocidade com que novas tecnologias são incorporadas ao nosso dia-a-dia.

* Endereço para correspondência: BioInfo - Estatística - IME-USP, Cep.05315-970, São Paulo, SP, Brasil.

Interessante é observar como o jovem consegue adaptar-se rapidamente ao uso das novas tecnologias sem ao menos se interessar pelo modo como são criadas ou desenvolvidas e em conhecer os conceitos usados na determinação de seus valores. O jovem torna-se rapidamente excelente usuário, superando em muito os adultos que foram educados sob um paradigma cultural totalmente distinto. O adulto de agora foi um jovem que questionava a necessidade de uma nova tecnologia disponível. Havia um desejo de entender-se como e por que aparatos tecnológicos eram criados e utilizados. A dinâmica atual parece não permitir que o jovem "desperdice" parte de seu tempo para questionar o que lhe está sendo oferecido em termos conceituais. Isto produz a idéia, talvez falsa, de o jovem só estar interessado por conhecimentos de uso imediato e, muitas vezes, de fácil aprendizado.

Como convencer um aluno de estatística de que análise, e não apenas cálculo, é fundamental para desenvolver seu raciocínio lógico? Em tempos idos, não era comum ao jovem questionar o professor sobre a utilidade do assunto que estava sendo ensinado. Um Professor era um Mestre! Hoje é um profissional do ensino, que é avaliado pela "produtividade". Por produtividade muitas vezes entende-se a quantidade, maior ou menor, de alunos que são aprovados. O mestre preocupava-se com a boa formação de seus alunos. O professor de hoje pode preocupar-se apenas com o mínimo que permita seu aluno avançar a uma nova etapa.

O Mercado de Trabalho, em função do perfil do novo profissional, também se transforma e hoje uma das exigências das empresas é o profissional com mestrado. Já não basta apenas um curso superior! O objetivo, como esperado, é a melhor qualificação do profissional. Provavelmente, em futuro próximo, irão exigir um Doutorado. Em resumo, hoje a pós-graduação já é parte da formação básica de um profissional. No passado, com a excelência da formação universitária, a pós-graduação era mais uma exigência acadêmica do que profissional.

Este trabalho, no lugar de criticar o jovem, questiona o ensino que lhe é oferecido. O objetivo é mostrar que a insegurança do jovem, ao discutir os fundamentos pelos quais um método é construído, pode ser consequência da forma imprecisa como os conceitos lhe são apresentados. Conceitos estatísticos, fundamentais aos métodos que

são usados no dia-a-dia do profissional, constituem o tema principal deste trabalho. A recomendação ao profissional é uma maior reflexão sobre seu instrumento de trabalho, os métodos estatísticos. O estudante deve sempre questionar seu professor quando um conceito não estiver claro. Esse questionamento deve ser sobre a lógica do método e não sobre seu uso imediato. Isto porque, ao entender completamente um método, o profissional saberá quando e como usá-lo. Imaginem as conseqüências da necessidade do uso imediato de um método quando o usuário não conhece a lógica de sua criação.

Este é um trabalho objetivo, pois toda a discussão é feita através de exemplos que mostram a inadequação de métodos de uso comum do estatístico. Evidentemente a inadequação diz respeito à situação específica do exemplo. O uso inquestionável de um determinado método pode trazer sérios problemas ao usuário. Muitas vezes a resposta a um problema pode não ser a resposta adequada. Conceitos imprecisos também aparecem como causa de soluções inadequadas. De fato, este texto questiona a falta de precisão dos professores de estatística quando ensinam certos conceitos fundamentais ao trabalho profissional. Não só a forma de ensinar, mas também os próprios conceitos precisam ser revistos. O risco da fuga de estudantes dos cursos de estatística irá aumentar caso não haja, por parte de nossos líderes, um esforço para uma revisão drástica no ensino da estatística.

Discutem-se na seqüência os conceitos de independência probabilística e estatística; os conceitos de dependência condicional e informação; os conceitos de intervalos de confiança e de credibilidade; e finalmente os conceitos de significância e evidência. Toda a discussão é feita por meio de exemplos especiais em contextos muito simples.

2. Independência: probabilística e estatística

Em estatística, o conceito de dependência está intimamente ligado ao conceito mais usual de associação. No entendimento comum, dois eventos estão associados quando um influencia a ocorrência do outro. Por exemplo, a ocorrência de um pode aumentar a chance de o outro ocorrer e neste caso dizemos que a associação é positiva.

Muitas vezes a ocorrência de um fenômeno pode diminuir ou mesmo eliminar a chance de o outro ocorrer e neste caso dizemos que a associação é negativa. Assim, a dependência entre duas quantidades aleatórias pode ocorrer em diferentes níveis, tanto positivamente quanto negativamente. Por outro lado, é sempre bom lembrar que os conceitos equivalentes de associação e dependência têm uma característica simétrica. Isto é, se x é dependente de y , com um determinado nível, então y é dependente de x com o mesmo nível.

A negação de dependência é a independência e este sim é um conceito preciso. Não há diferentes níveis de independência! Afirma-se que duas quantidades aleatórias são independentes quando, ao conhecer-se o valor de uma, a distribuição da outra não se altera. Em outras palavras, se x e y são duas quantidades aleatórias independentes então $p(y) = p(y | x)$; isto é, o fato de x ter sido observado não altera a distribuição de y . Um bom exercício para o leitor é provar que as seguintes três igualdades são equivalentes:

$$\text{i) } p(y) = p(y | x);$$

$$\text{ii) } p(x) = p(x | y); \text{ e}$$

$$\text{iii) } p(x, y) = p(y)p(x).$$

O conceito de independência probabilística é naturalmente estendido para o de independência condicional quando outra quantidade z entra no contexto. As quantidades x e y são ditas condicionalmente independentes, dado z , se

$$p(x, y | z) = p(y | z)p(x | z).$$

Outro bom exercício para o leitor é mostrar a equivalência das seguintes igualdades:

$$\text{iv) } p(y | z) = p(y | x, z);$$

$$\text{v) } p(x | y) = p(x | y, z); \text{ e}$$

$$\text{vi) } p(x, y | z) = p(y | z)p(x | z).$$

Independência estatística é na verdade independência condicional. Em estatística o parâmetro, desconhecido e de interesse, exerce o mesmo papel da quantidade z . Os Exemplos 1, 2 e 3 ilustram os conceitos de independência probabilística e estatística.

Exemplo 1: Normal padrão

Considerem-se duas variáveis aleatórias normais independentes, x e y , com mesma distribuição normal com média zero e variância um. Inicialmente, sabe-se que embora os valores das duas quantidades possam ocorrer em toda a reta, muito provavelmente não estarão fora do intervalo $[-6;+6]$. Além disso, a informação de que $x = 3$, por exemplo, não irá alterar a distribuição de y . Isto é, $p(y) = p(y|x)$. Novamente y pertence com grande chance ao intervalo $[-6;+6]$. O conhecimento da ocorrência de $x = 3$ não alterou a distribuição de probabilidades inicialmente proposta para y . No presente caso, a independência é a probabilística no sentido usual.

Exemplo 2: Normal com média desconhecida

Sejam duas variáveis aleatórias normais independentes, x e y , com a mesma distribuição normal com média z , desconhecida, e variância um. Como o valor da média, z , é desconhecido, tanto x como y podem variar em toda a reta, sem preferências. A observação $x = 3$ agora é bastante informativa com respeito à distribuição de y . Evidentemente, existe uma grande chance de que $x = 3$ esteja no intervalo $[z-6; z+6]$. Assim, $-3 < z < 9$ e, portanto, y muito provavelmente irá pertencer ao intervalo $[-9;15]$. Note-se que, antes da observação de x , o valor de y era completamente desconhecido e poderia assumir qualquer valor em toda a reta. No entanto, após observar-se $x = 3$, sabe-se que quase certamente y pertence ao intervalo finito $[-9;15]$. Isto prova que x e y não podem ser independentes, visto que a observação do valor de x modifica de fato a distribuição de y . No exemplo acima o correto seria dizer: caso z fosse conhecido, x e y seriam independentes. Esse é o caso da independência estatística. Na linguagem probabilística, ao se considerar z como variável aleatória, x e

y seriam condicionalmente (em relação a z) independentes. Na perspectiva freqüentista da estatística, o parâmetro, representado aqui por z , não é considerado como variável aleatória. Conseqüentemente, sem uma estrutura probabilística para z , o uso do conceito de independência condicional estaria prejudicado. Contudo, o estatístico poderia sempre afirmar que, no caso do valor de z ser conhecido, fosse ele qual fosse, x e y seriam independentes. Isto é,

$$p(x, y | z) = p(y | z)p(x | z).$$

Para, de forma definitiva, sensibilizar os professores de estatística, o Exemplo 3 é mais dramático do que o anterior.

Exemplo 3: Uniforme com média desconhecida

Considerem-se duas variáveis aleatórias independentes, x e y , com mesma distribuição uniforme no intervalo $[z-1; z+1]$, onde a média z é desconhecida. No início, pode-se dizer que $-\infty < y < +\infty$. No entanto, observando-se $x=3$, conclui-se que $2 < z < 4$ e, assim, y certamente pertencerá ao intervalo $[1; 5]$. Isto é, o conhecimento de que $x=3$ faz com que, novamente, apresente-se o caso de independência condicional (estatística) entre x e y , dado z .

Esse terceiro exemplo mostra que, embora ocorra a independência estatística (correspondendo, no caso de estrutura probabilística para z , à independência condicional de x e y , dado z), na verdade tem-se uma forte dependência (não-condicional). Note-se que ambas as variáveis possuem um elemento desconhecido em comum nas suas distribuições e esse elemento, o parâmetro z , é o que provoca a forte associação entre x e y .

3. Informação e dependência condicional

O exemplo discutido nesta seção é de igual importância, pois ressalta mais uma vez a diferença entre o trabalho estatístico e o probabilístico. O leitor deve entender, no entanto, que um não exclui o outro. Na verdade, o probabilístico complementa o estatístico, e vice-versa!

Apresenta-se apenas um exemplo, suficiente para a discussão, por ser crucial para o ensino da Estatística. Lembre-se de que a função de verossimilhança carrega toda a informação, contida nos dados x , sobre os parâmetros z . Diferenciem-se bem os dois elementos do modelo estatístico: (a) a distribuição amostral é caracterizada pela função de probabilidade de x para cada valor fixado, z_0 , do parâmetro z , isto é, $p(x|z_0)$; e (b) a função de verossimilhança para a observação x_0 de x é a distribuição amostral avaliada em x_0 , como função do parâmetro z , isto é, $L(z|x_0) = p(x_0|z)$. Assim, a distribuição amostral pode ser caracterizada pela classe dos conjuntos $\mathfrak{F}_z = \{f(x|z); x \in X\}$ e a verossimilhança da observação x pelo conjunto $\mathfrak{F}_x = \{f(x|z); z \in Z\}$ onde X e Z são, respectivamente, os espaços amostral e paramétrico.

Exemplo 4: Um par de filhos

A escola de dança do bairro compete todos os anos com um grupo de 10 meninas no concurso nacional de artes. A professora Lily toma conhecimento de que as duas irmãs que fazem parte do grupo vão acompanhar os pais na imigração para o Chile. Ela é informada de que a nova família, que vai morar na casa onde moram atualmente as duas bailarinas, possui um par de crianças. Lily, então, pondera que se forem duas meninas poderão substituir as duas que estão saindo do grupo. Dessa forma, terá tempo de treinar as novas componentes para a próxima competição. Acha, então, que tem probabilidade $\frac{1}{4}$ de o par ser formado por duas meninas. Seu espaço de probabilidades equiprováveis é $\{(m,m);(m,h);(h,m);(h,h)\}$. Conversando, então, com seu irmão, Jony, corretor de imóveis, a professora é informada de que o par é formado por pelo menos uma menina. Essa informação Jony obteve quando ouviu por telefone a mãe dizendo

“fique quieta minha filha” para uma criança que fazia algum ruído perto do telefone. Lily fica alegre já que sua probabilidade aumentou para $1/3$; desse modo seu espaço de probabilidades equiprováveis é agora $\{(m, m); (m, h); (h, m)\}$. Mary, a esposa de Jony, entra na sala e diz que uma das crianças da nova moradora é uma menina que está no carro. Lily sai e vai ao carro para ver a criança. Fica mais feliz, pois entende que sua probabilidade aumenta para $1/2$ após ver a menina e se perguntar “qual a probabilidade de a outra criança ser menina?”. Seu espaço de probabilidades equiprováveis, depois de ver a criança, é $\{m; h\}$, visto que a única fonte de incerteza é o sexo da outra criança.

Aceitar passivamente o que foi apresentado na dramatização acima, implica aceitar-se que a afirmação “pelo menos uma menina no par” possui diferentes pesos quando recebida por diferentes canais: audição e visão.

O olhar de um estatístico permite incorporar outras informações ocorridas no processo de coleta de dados. Para entender a racionalidade estatística, é preciso definir apropriadamente todas as entidades do processo. Seja o parâmetro $z = 1$ no caso de um par de meninas e $z = 0$ no caso de um casal. A probabilidade *a priori* de z é definida por $p(1) = 1/3 = 1 - p(0)$. Como variável observacional, para definir a verossimilhança, considere-se $x = 1$ se a criança que acompanha Mary é uma menina e $x = 0$ se menino. A função de verossimilhança, $p(x | z) = L(z | x)$, no caso presente de $x = 1$ é $L(z | 1)$. No caso de $z = 1$ (par de meninas), $L(1 | 1) = 1$ e, se $z = 0$ (um casal), $L(0 | 1) = q$, onde q é um valor do intervalo $[0; 1]$. Equivalentemente, pode-se escrever $p(x = 1 | z = 0) = q$ e $p(x = 1 | z = 1) = 1$. Com um cálculo simples e após lembrar que sua observação foi $x = 1$, o leitor deve ser capaz de mostrar que a probabilidade *a posteriori* de $z = 1$ é, então,

$$p(z = 1 | x = 1) = [1 + 2q]^{-1}.$$

A princípio Lily pensou que Mary teria probabilidades iguais ($q = 1/2$) de trazer menino ou menina, no caso de o par ser um casal de crianças ($z = 0$). Neste caso, Lily estaria certa em se alegrar com o aumento da sua probabilidade para um par de

meninas, pois, se $q = 1/2$, a probabilidade *a posteriori* de $z = 1$ seria $1/2$. Mas Lily logo percebeu seu erro! Mary levou a criança, a pedido da mãe ocupada, para provar um vestidinho na nova loja de roupas femininas infantis, ali em frente. Assim Lily teve de considerar $q = 1$ e nesse caso *a posteriori* seria $1/3$, o mesmo valor da probabilidade *a priori* após a informação transmitida por Jony. Nesse caso a verossimilhança é constante e assim não informativa!

Como conclusão desta seção, o leitor deve entender que no momento da definição do modelo de probabilidades, é necessário sempre levar em consideração o modo como a ocorrência de eventos é observada.

4 . Confiança e credibilidade

O exemplo discutido nesta seção diz respeito à estimação por intervalo. O conceito de intervalo de confiança é o que mais dificulta a vida do estatístico. Muitas vezes este precisa explicar ao cientista que o intervalo numérico apresentado, acompanhado da porcentagem 95%, não quer dizer que a probabilidade de esse intervalo conter o verdadeiro valor do parâmetro é 95%. Deve na verdade explicar que, se o procedimento pudesse ser repetido uma série de vezes, os intervalos construídos conteriam o verdadeiro valor do parâmetro em 95% das vezes. Para se entender melhor a dificuldade que esse procedimento representa, o exemplo, a seguir, mostra a diferença entre probabilidade e confiança.

Exemplo 5: Uniforme com média desconhecida

Considerem-se quatro observações, $\{u, v, x, y\}$, independentes (estatisticamente) e identicamente distribuídas segundo uma uniforme no intervalo $[z - 1/2; z + 1/2]$. O objetivo é construir um intervalo de confiança para z a partir das quatro observações. Represente-se por m e M , respectivamente, o menor e o maior valores amostrais. Não é difícil ver que o intervalo $[m; M]$ é um intervalo com confiança 87,5%, pois $1 - 2(1/2)^4 = 0,875$. Lembre-se de que $(1/2)^4$ é a probabilidade de que as quatro

observações estejam acima (ou abaixo) da média z ; isto é,

$$p(m > z) = p(M < z) = (1/2)^4 = 0,0625.$$

Os quatro valores efetivamente observados foram: 1,11; 1,27; 1,43; 1,59. Um estatístico clássico concluiria que [1,11; 1,59] é um intervalo com 87,5% de confiança. Note-se que, se no lugar de 1,59 tivéssemos 1,50 ou mesmo 1,62, a confiança seria a mesma, embora os comprimentos dos intervalos fossem, respectivamente, 0,48, 0,39 e 0,51. Mesmo no caso em que M fosse 1,91 a confiança do intervalo [1,11;1,91] continuaria sendo 87,5%. O que um estatístico menos comprometido com métodos faria? Certamente consideraria que $M < z + 0,5$ e $m > z - 0,5$. Isto é, o mínimo e o máximo de qualquer amostra devem situar-se no intervalo de variação da variável que está sendo estudada. Com isso seria possível dizer que com probabilidade 1, certeza total, $M - 0,5 < z < m + 0,5$. Isto é, com certeza, o intervalo $[M - 0,5; m + 0,5]$ contém z ! Resumindo: se $m = 1,11$, os casos em que $M = 1,59, M = 1,50, M = 1,62$ e $M = 1,91$ produziriam, respectivamente, os intervalos [1,09;1,61], [1,00;1,61], [1,12;1,61] e [1,41;1,61], cujos comprimentos seriam 0,52, 0,61, 0,49 e 0,21. Todos esses intervalos, com certeza – com probabilidade 1 –, conteriam o verdadeiro valor do parâmetro z . Assim, o resultado amostral efetivamente observado pode ser mais ou menos informativo. Isto é, quanto maior a diferença $M - m$, mais informativo deve ser o resultado da inferência sobre o valor de z . Note-se que, com o procedimento do clássico intervalo de confiança de 87,5%, os intervalos seriam [1,11;1,59], [1,11;1,50], [1,11;1,62] e [1,11;1,91]. Os comprimentos dos intervalos seriam 0,48, 0,39, 0,51 e 0,81. A conclusão é que o procedimento apresentado oferece menos informação ao cliente – intervalo de grande comprimento – justamente quando a amostra é mais informativa, M e m distantes. Isto é uma incoerência!

Para uma solução Bayesiana do problema estudado no Exemplo 5, considere-se *a priori* que o parâmetro z se distribui uniformemente no intervalo [-1 000;1 000]. Com

respeito às amostras descritas acima, as distribuições *a posteriori* seriam uniformes nos intervalos [1,09;1,61], [1,00;1,61], [1,12;1,61] e [1,41;1,61]. De posse dessas distribuições, poder-se-iam construir intervalos centrais com 0,875 de probabilidade. Esses intervalos são conhecidos como intervalos de credibilidade com 87,5% de credibilidade. Conseqüentemente, [1,12;1,58], [1,04;1,57], [1,15;1,58] e [1,42;1,60] seriam os intervalos com credibilidade 87,5% e comprimentos 0,46, 0,53, 0,43 e 0,18. Observe-se a coerência desse tipo de inferência em que amostras mais informativas oferecem respostas superiores; ao escrevermos 87,5% de credibilidade estamos, como é desejável, falando de o intervalo conter o valor desconhecido do parâmetro, z , com probabilidade 0,875.

A seção 5 conclui este artigo com uma discussão sobre o método estatístico “teste de significância”, o mais contemplado nas diversas áreas de aplicação. Novamente, o interesse é mostrar que certos cuidados devem ser tomados e os resultados corretamente interpretados quando um determinado método é utilizado.

5. Significância e evidência

Neste capítulo discute-se, por meio de dois exemplos, o conceito de significância. Foram usadas as recomendações comuns para o cálculo dos valores- p em problemas de teste de significância. Trata-se de dois exemplos simples e talvez comuns ao dia-a-dia de um estatístico.

É comum aos cursos de estatística o fato de não ser dada ênfase à hipótese alternativa quando testes de hipóteses baseiam-se no conceito de significância. Após a obtenção das observações amostrais, apenas a hipótese nula é usada no momento da definição da distribuição nula e suas caudas. Muitas vezes não se considera o fato de distribuições sob as hipóteses alternativas possuírem caudas menos pesadas do que aquelas sob a hipótese nula. Em alguns casos, os níveis de significância não devem ser áreas das caudas, mas sim áreas centrais da distribuição nula. Os exemplos discutidos, a seguir, mostram quão importantes podem ser as hipóteses alternativas na definição da regra aceita/rejeita.

Exemplo 6: Bolinhas na urna

Apresenta-se uma urna com três bolinhas de diferentes cores: preta, verde e branca. Após três retiradas de uma bola da urna, anotaram-se as quantidades $x =$ número de pretas (0 ou 1) e $y =$ número de verdes (0 ou 1). Observado o resultado $(x,y)=(1,1)$, o interesse é testar a hipótese H : a amostra foi obtida com reposição, contra a alternativa A : a amostra foi obtida sem reposição. A Tabela 1 apresenta a distribuição de (x,y) , multiplicada por 27, quando H é verdadeira. Note-se que o valor- p é igual a 100%, se a definição da cauda incluir o observado $(1,1)$, ou 77,8%, no caso de sua exclusão. Seguindo-se as regras ensinadas em cursos de estatística, H não seria rejeitada e A não seria considerada. Isto é, se $(x,y) = (1,1)$, H é aceita! A questão é: A pode ser desconsiderada nesse caso? Observe-se que amostras que teriam probabilidade de ocorrência baixa seriam aquelas diferentes de $(1,1)$. Entretanto, qualquer amostra diferente de $(1,1)$ teria probabilidade zero de ocorrer no caso de A verdadeira. Isto é, os pontos que poderiam levar à rejeição de H seriam pontos impossíveis de ocorrer sob A . Novamente, observa-se uma incoerência!

Tabela 1: Distribuição nula de (x,y) do Exemplo 6

y	0	1	2	3
x				
0	1	3	3	1
1	3	6	3	0
2	3	3	0	0
3	1	0	0	0

Analisando-se a questão com os olhos de um estatístico preocupado com o problema a ser resolvido e não com um método disponível, a razão de verossimilhanças seria algo a ser considerada. A razão de verossimilhanças para um resultado (x,y) é a probabilidade desse resultado, quando A é verdadeira, dividido pela probabilidade,

quando H é verdadeira. Isto é, $r = (x, y) = p(x, y | A) \div p(x, y | H)$. Para todos os resultados possíveis, exceto o ponto (1,1), o valor de r seria zero. No caso do resultado observado, $r(1, 1) = 9/2 = 4,5$. Isto prova que privilegiar A em detrimento de H é a conclusão mais acertada com o resultado $(x, y) = (1, 1)$. Qualquer outro resultado indicaria a aceitação de H .

Sob a perspectiva Bayesiana a análise fica mais simples. Supondo-se equiprobabilidades *a priori*, $p(H) = p(A) = 1/2$, a conseqüente distribuição *a posteriori* de A seria $p(A | 1, 1) = 1 - p(H | 1, 1) = 1 - 6/33 = 9/11$. No caso de qualquer outro resultado, H certamente seria verdadeira. Assim, com probabilidade *a priori* considerada, a observação de $(x, y) = (1, 1)$ tem como conseqüência o fato de a probabilidade de A passar a ser quatro vezes maior do que a probabilidade de H . Isto é, a amostra observada favorece A em detrimento de H .

O exemplo final deste manuscrito é mais comum ao professor de estatística, pois se refere à distribuição normal, a mais discutida em salas de aula.

Exemplo 7: Normal com variância desconhecida

Considerem-se três observações, x , y e z , estatisticamente independentes e identicamente distribuídas segundo uma normal com média zero e variância desconhecida, representada por v , $N(0, v)$. A estatística suficiente neste caso é a soma dos quadrados das observações, $s = x^2 + y^2 + z^2$. As hipóteses a serem confrontadas são a nula $H: v = 2$ contra a alternativa $A: v \neq 2$. É importante ressaltar-se que a variável aleatória s/v é distribuída segundo uma χ^2 com 3 graus de liberdade (gl), independente do valor que v pode assumir. Considere-se que os valores observados sejam $x = 0,6$, $y = 0,3$ e $z = -0,2$, conseqüentemente, $s = 0,49$. Sob a hipótese H , $N(0, 2)$ é a distribuição nula e suas caudas, delimitadas por essas observações, seriam $C(x) = 89\%$, $C(y) = 83\%$ e $C(z) = 67\%$. Multiplicando-se as áreas das três caudas, o resultado seria $C(x, y, z) \cong 50\%$. Por outro lado, a correspondente área da cauda da

distribuição nula (χ^2 com $gl = 3$) de $s/2$ é calculada como

$$p(s > 0,49 | H) = p(\chi^2 > 0,245) = 0,97.$$

Isto é, a amostra $(x, y, z) = (0,6; 0,3; -0,2)$ produz uma significância (valor- p) de 97%, favorecendo H contra A . Todos esses cálculos padrões privilegiam a hipótese H , em detrimento de A . Outra forma de obter-se um valor- p seria concentrar as atenções apenas em s , abandonando a amostra original, visto que s é uma estatística suficiente para (x, y, z) com respeito ao parâmetro z . Ao considerar-se que valor da significância é a área das caudas da distribuição nula, o primeiro passo seria a definição das caudas da distribuição nula de $s/2$, de acordo com seu valor obtido na amostra, 0,245. O ponto com mesma densidade na distribuição χ^2 com $gl = 3$ é 2,6112. Com as caudas definidas pelos limites (de igual densidade) 0,245 e 2,6112, o novo valor- p seria

$$p(\chi^2 < 0,245) + p(\chi^2 > 2,6112) = 0,02998 + 0,45553 = 0,4855.$$

Dessa forma, a significância é 48,55%, novamente privilegiando H contra A .

Usando-se métodos padrões, como os descritos acima, as observações encontradas favoreceriam H contra A . Contudo, usando a intuição, não é absurdo concluir-se que a observação $s = 0,49$ deveria privilegiar valores de v inferiores a 2: As três observações pertencem a um intervalo contido em $(-1; 1)$. Relevante seria o leitor entender que valores baixos de s correspondem a pontos amostrais centrais. Tais pontos são mais prováveis de ocorrer – maior densidade – quando v for baixo, como, por exemplo, $v \leq 1$. Neste caso, a hipótese alternativa é que deveria ser privilegiada e não a nula. A conclusão desta discussão indicaria que as regiões críticas do teste seriam aquelas que contivessem valores amostrais próximos de zero (caso da amostra estudada) ou bem distantes de zero, isto é nas caudas da distribuição nula. O problema aqui é a definição de região crítica: apenas a hipótese nula é usada na delimitação das caudas, usadas como região crítica. A seguir, ambas as hipóteses serão usadas na

definição de regiões que beneficiam A , em detrimento de H , mais do que a amostra efetivamente observada.

Para uma visão menos objetiva do Exemplo 7, considere-se que, sob A , $u = v^{-1}$ tem uma distribuição *a priori* χ^2 com um grau de liberdade. Ao usar-se a função de densidade $h(u)$ dessa distribuição como um sistema de preferências no espaço (paramétrico) dos possíveis valores de u , a densidade preditiva sob A , $f(s|A)$, é definida como a média – ponderada por $h(u)$ – da densidade de s , $f(s|u)$. Obtém-se assim,

$$f(s|A) = \int h(u)f(s|u) du = \left(\frac{2}{\pi}\right)(\sqrt{s})(s+1)^{-3}.$$

Com as densidades nula e preditiva – $f(s|H) = f(s|v=2)$ e $f(s|A)$ – define-se a razão de verossimilhanças ou Fator de Bayes:

$$r(s) = f(s|A) \div f(s|H) = \left(\frac{8}{\sqrt{\pi}}\right) \left[(s+1)^{-2}\right] \exp\left(\frac{s}{4}\right).$$

O fator de Bayes pode ser usado como um índice de ordem no espaço amostral. Este índice, $r(s)$, permite o cálculo de um valor- p que leva em consideração ambas as hipóteses. Com a observação $s = 0,49$, procura-se o outro ponto s' com mesmo valor de r . No caso do exemplo, o ponto $s' = 22,5845$ satisfaz essa propriedade; isto é, $r(0,49) = r(22,5845) = 4,073$. Com esses pontos delimitando as caudas da distribuição nula, obtém-se o valor- p modificado, $pv = 4\%$. Isto é,

$$1 - p(0,49 < s < 22,5845 | H) = 1 - p(0,245 < \chi^2 < 11,292) = 0,04.$$

Seguindo a intuição do estatístico competente, este valor de pv , se considerado como *índice de significância*, privilegia a hipótese alternativa em detrimento da hipótese nula. Por outro lado, se as probabilidades *a priori* das duas hipóteses fossem

consideradas equiprováveis, a probabilidade *a posteriori* seria

$$p(\mathbf{H} | s) = (1 + 4,073)^{-1} = 0,197,$$

novamente favorecendo A contra H .

Concentrando a atenção apenas na função de verossimilhança, que carrega toda a informação contida na amostra sobre o parâmetro, o estatístico na maioria das vezes tira as conclusões corretas e de forma coerente. Com a parametrização considerada, a hipótese nula é rescrita como $\mathbf{H} : u = v^{-1} = 1/2$. A função de verossimilhança de u é proporcional a uma densidade gama com parâmetros $A = 5/2$ e $B = 0,245$. Por outro lado, os valores assumidos por esta função nos pontos $u = 1/2$ e $u = 24,2696$ são iguais. Conseqüentemente, todo valor de u no intervalo $(0,5; 24,2696)$ é mais verossímil do que $1/2$, o valor de u da hipótese nula. O conjunto T formado por todos os pontos paramétricos mais verossímeis do que o da hipótese nula é denominado *conjunto tangente*: $T = (0,5; 24,2696)$ no presente caso. Como a função $g(u)$, a função de verossimilhança normalizada por sua integral, é a densidade Gama com parâmetros $A = 5/2$ e $B = 0,245$, a área sob a função de verossimilhança em T é igual a 96,23% da área total. Isto é um indicativo de que há uma predominância de pontos mais verossímeis do que a hipótese nula. Como uma alternativa ao valor- p , define-se o valor da *evidência* da amostra em favor de H , valor- e ou ev , como o complemento da área do conjunto tangente; isto é, $ev(t) = 3,77\%$. Novamente, como esperado, apresenta-se um indicativo contra a hipótese nula em favor da alternativa.

Esta última avaliação da amostra, com relação à hipótese nula, foi baseada exclusivamente na função de verossimilhança, em completa concordância com o princípio da verossimilhança. De acordo com esse princípio, duas funções proporcionais de verossimilhança devem produzir a mesma inferência estatística. Evidentemente, ev pode ser calculado sempre que uma densidade de probabilidade esteja sendo definida no espaço paramétrico, como é o caso de uma densidade *a posteriori*, ou mesmo de uma função de verossimilhança cuja integral seja finita. No caso discutido acima, os

valores- e são baseados nas caudas da verossimilhança, caudas essas definidas a partir da hipótese nula. Uma das vantagens deste método em relação a outros é que não existe restrição ao número de parâmetros envolvidos. Além disso, não há a necessidade de eliminação de parâmetros *indesejáveis*, uma prática comum ao estatístico que atua sob o paradigma freqüentista. Não há, teoricamente, restrições de dimensão nem ao espaço paramétrico nem ao amostral. Evidentemente, quanto maior a dimensão, maior o desafio computacional.

Embora pv e ev sejam probabilidades calculadas, respectivamente, no espaço amostral e no espaço paramétrico, seus valores apontam para a mesma conclusão: evidência contra H favorecendo A . As conclusões obtidas com os três métodos são:

1. ao usar-se a receita freqüentista de olhar apenas a densidade sob a hipótese nula, desconsiderando a hipótese alternativa, os dados observados favorecem H em detrimento de A ;
2. ao considerar-se a hipótese alternativa no cálculo do valor- p modificado, os mesmos dados desfavorecem H em favor de A ; e
3. sob uma visão genuinamente Bayesiana, onde a única contribuição da amostra é a função de verossimilhança, a conclusão baseada no valor da evidência também favorece A em detrimento de H .

Resumindo: as significâncias de H , $pv \cong ev \cong 4\%$, obtidas com os dois métodos, clássico modificado e Bayesiano, indicam a rejeição de H , favorecendo A , enquanto o método padrão, com o valor- p de 49%, aponta para a direção inversa.

6. Ilustração gráfica

Nesta seção o leitor encontra uma série de ilustrações que visam a auxiliá-lo no entendimento dos exemplos apresentados. Densidades normais e uniformes, discutidas nos exemplos 2 e 3, são apresentadas na Figura 1. A Figura 2 apresenta os valores da probabilidade *a posteriori* do Exemplo 4 em função das possíveis verossimilhanças. O objetivo principal é ressaltar que a subjetividade pode estar, muitas vezes, na licitação da função de verossimilhança e não apenas da distribuição *a priori*.

O Exemplo 5 apresentou o caso onde intervalos de confiança efetivamente observados produzem uma incoerência de ordem prática. Quanto melhor a amostra for, pior é a precisão produzida pelo método. Em outras palavras, o intervalo de confiança aumenta seu comprimento, diminuindo a precisão, se uma nova amostra for mais informativa - o máximo M mais distante do mínimo m . O método de intervalos de credibilidade discutido, caminha na direção oposta, isto é quanto mais informativa a amostra, maior a precisão obtida. A Figura 3 apresenta os dois tipos de intervalos para as quatro amostras discutidas. Com a amostra 4, a mais informativa, a incoerência do método padrão fica evidente. Com uma amostra ainda mais informativa, por exemplo se $m \cong 1,1$ e $M \cong 2,1$, qualquer pessoa deveria concluir que certamente $z \cong 1,55$. Mas, o intervalo com confiança de 87,5% seria $[1,1; 2,1]$, produzindo uma inferência quase totalmente não informativa.

Como na Figura 1, a Figura 4 apresenta possíveis densidades normais para x , pertencentes ao conjunto de densidades que formam o modelo amostral de $x|u$. Note-se que todas as três observações possuem maior densidade quando $u = 1$, um ponto da hipótese alternativa. Como os procedimentos de inferência são obtidos a partir da estatística suficiente s , a Figura 5 apresenta alternativas das densidades amostrais de $s|u$.

A Figura 6 apresenta a função de verossimilhança normalizada por sua área. As caudas e o conjunto tangente são ressaltados. O conjunto tangente é definido em relação à hipótese $H: u = 1/2$.

A Figura 7 apresenta a razão de verossimilhança $r(x)$ para amostras unitárias. Finalmente, para amostras de três observações, a Figura 8 é a razão de verossimilhanças $r(s)$ para a estatística suficiente s .

Figura 1- Densidades normais e uniformes

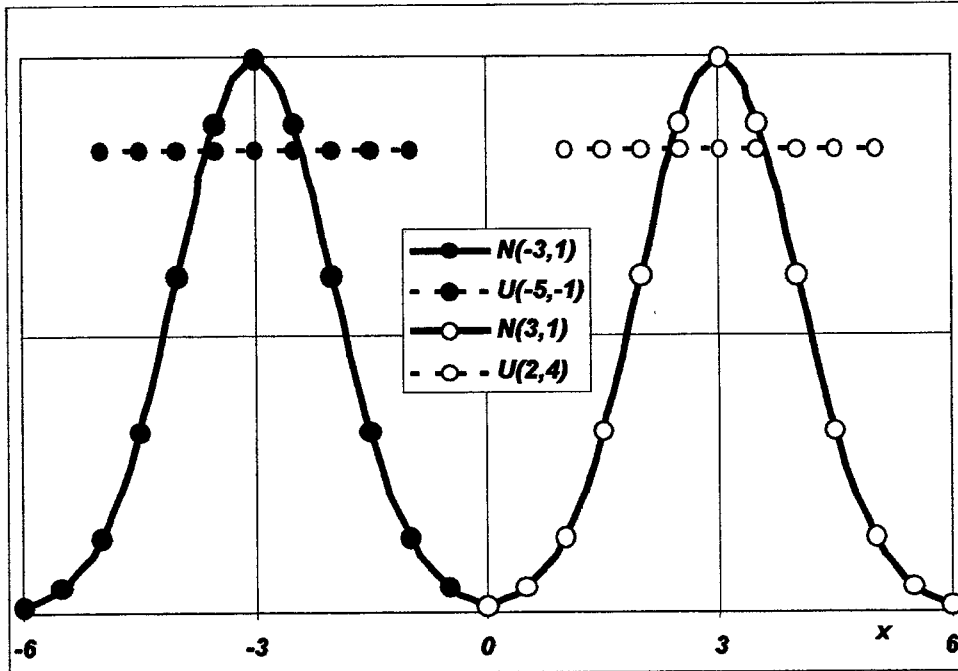


Figura 2 - Probabilidade a Posteriori em Função da Função Verossimilhança (q): $(1+2q)^{-1}$

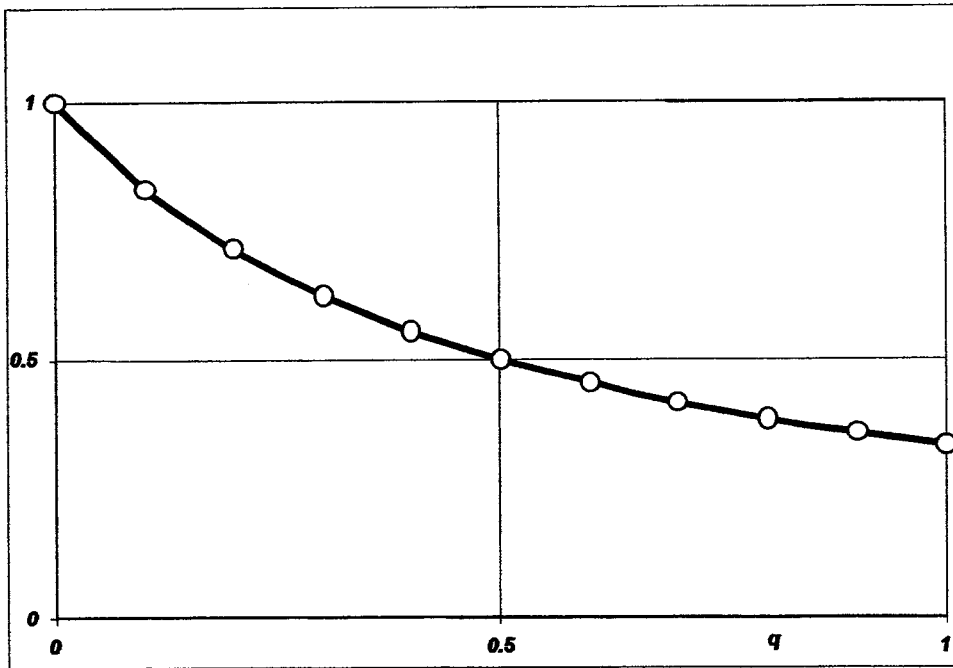


Figura 3 - Comparação de Intervalos Clássicos e Bayesianos:
87,5% de Confiança ou de Credibilidade

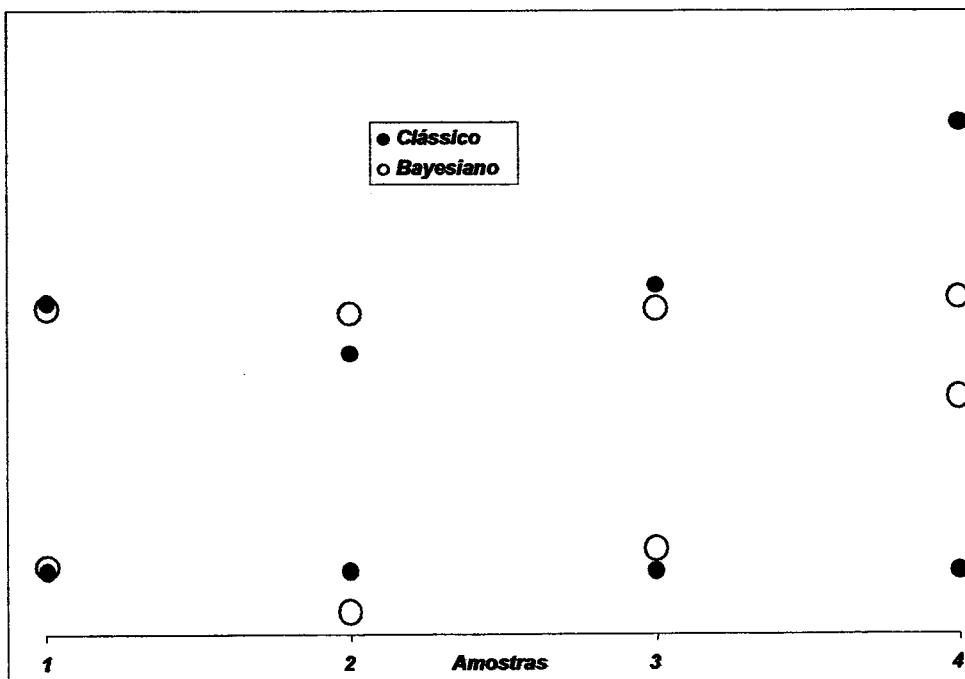


Figura 4 - Densidades para $x | u$

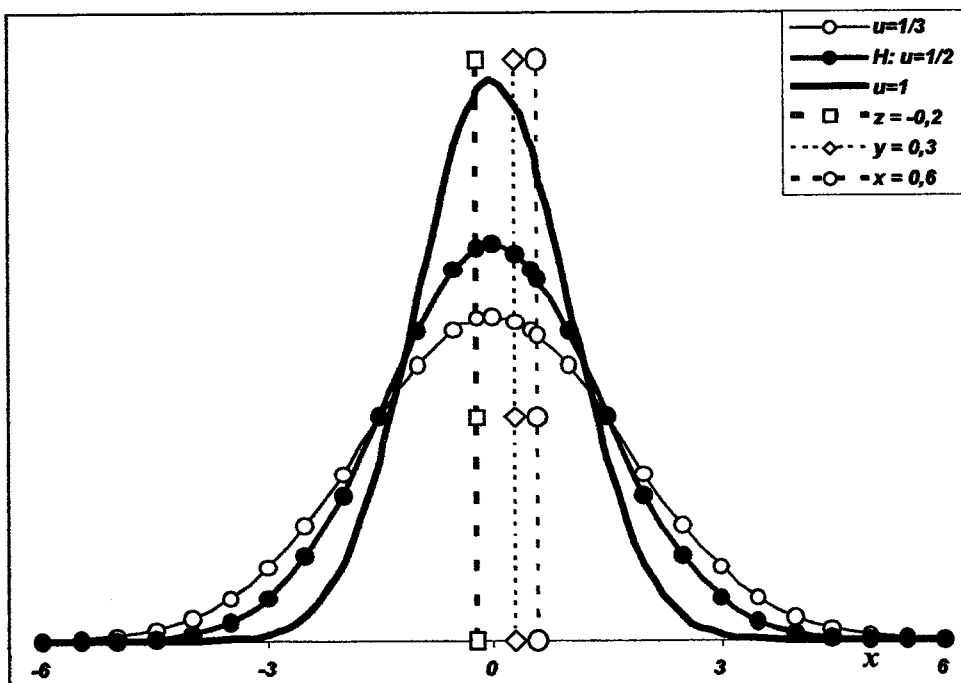


Figura 5 - Densidades para $S | U$

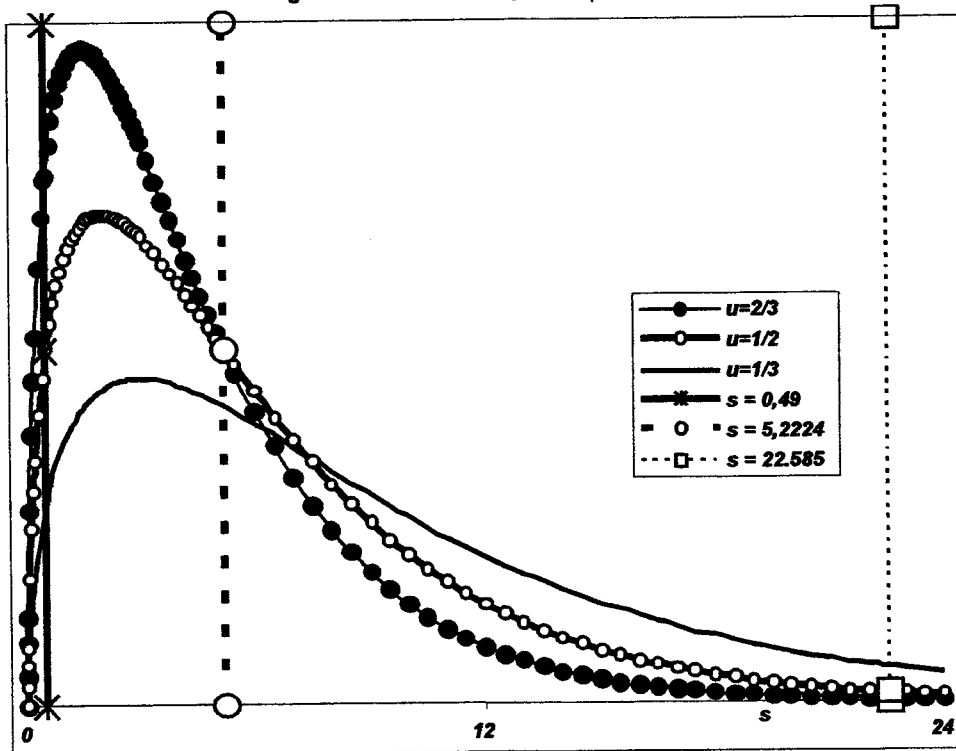


Figura 6 - Função de verossimilhança do ponto $s = 0,49$: $T = [0,5; 24,2696]$ e $ev = 0,04$

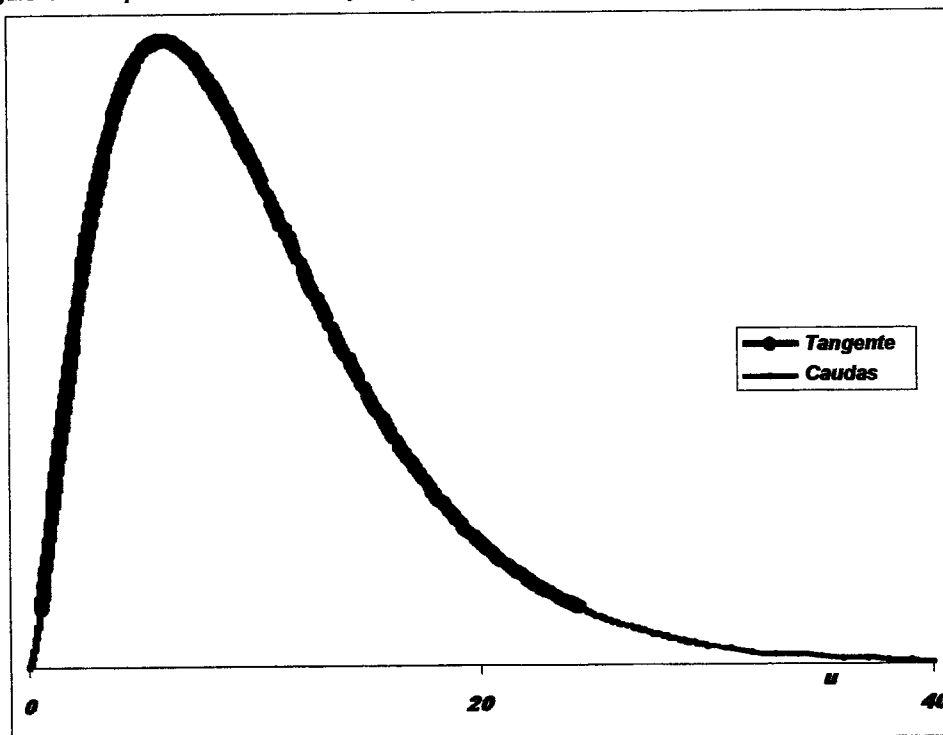


Figura 7 – Fator de Bayes para uma observação x

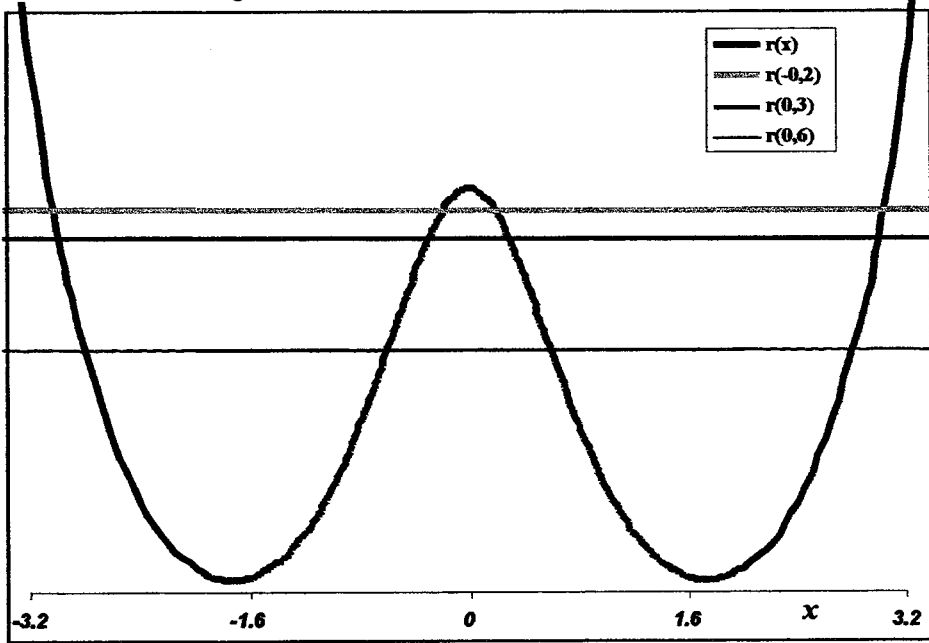
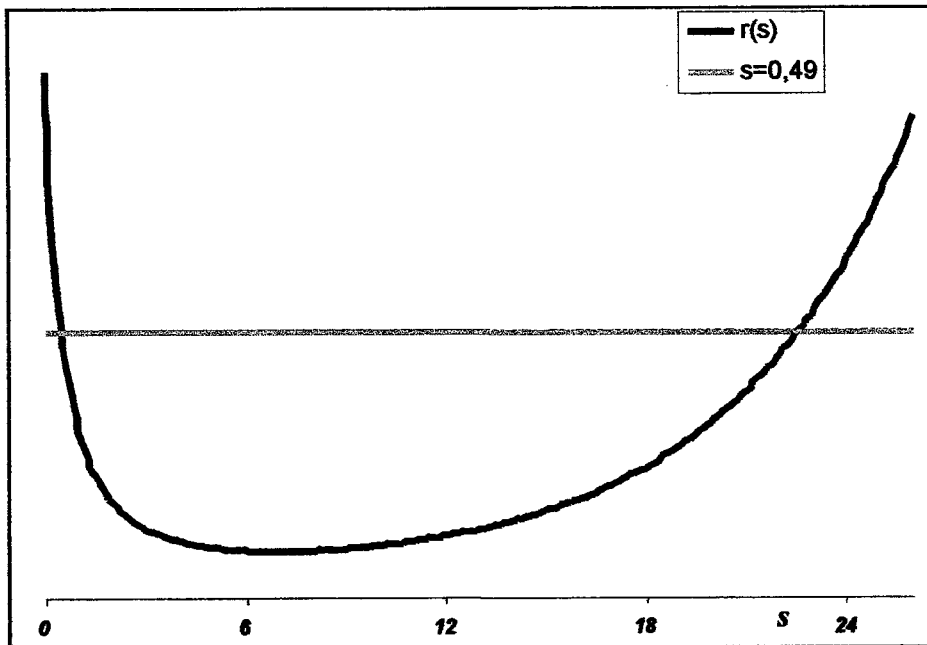


Figura 8 - Fator de Bayes para s



7. Considerações finais

Este trabalho nasceu de desafios e incentivos que o autor recebeu de seus colegas. Encontros informais, durante a participação em bancas ou seminários, permitiram ao autor expor idéias e discussões que aparentemente surpreendiam os presentes.

O professor Dev Basu, supervisor do autor durante o programa de doutorado, foi o mestre dos contra-exemplos. Muitos dos seus artigos tinham a perspectiva do desafio intelectual sobre o entendimento dos fundamentos da estatística. Dev e Kemp - Oscar Kempthorne outro ilustre professor - criavam um ambiente de conflito que muito auxiliou o desenvolvimento da estatística. A impressão é a de que não concordavam entre si, mas certamente se desafiavam mutuamente para o desenvolvimento das idéias que tão brilhantemente defenderam. Interessante é ressaltar que a admiração do autor pelos professores Dev e Kemp, bem como por I. J. Good e Bruno De Finetti, talvez se deva ao fato de todos, sem exceção, terem feito história por suas controvérsias e pela apresentação de alternativas metodológicas coerentes.

O presente artigo é a tentativa do autor de criar, no ambiente da estatística brasileira, o conflito de idéias para impedir que métodos sejam usados apenas porque foram criados por líderes de renome. É evidente que o artigo é pretensioso! Mas é justamente essa pretensão que talvez leve os colegas ao conflito. O desafio é fazer com que todos saibam defender, com conhecimento dos fundamentos, os métodos que usam.

Os agradecimentos vão para os colegas de departamento por expor o autor aos desafios que surgem com problemas práticos que se apresentam no Centro de Estatística Aplicada - CEA. O principal responsável, entretanto, é o Professor Francisco Louzada-Neto, do departamento de estatística da UFSCar, por insistir com o autor sobre a necessidade da divulgação de idéias em forma escrita.

Referências bibliográficas

- RICHARD BARLOW (1998), *Engineering Reliability*, SIAM.
- DEV BASU (1988), *Statistical Information & Likelihood: A Collection of Critical Essays*, JK Gosh editor, Springer-Verlag. Lecture Notes in Statistics #45.
- DAVID BLACKWELL (1969), *Basic Statistics*, McGraw-Hill.
- BRUNO DE FINETTI (1972), *Probability, induction, and statistics*, Wiley.
- MORRIS DEGROOT (1986), *Probability & Statistics*, Adison-Wesley
- IRVING J. GOOD (1983), *Good Thinking*, U. Minnesota Press.
- OSCAR KEMPTHORNE & LEROY FOLKS (1971), *Probability, Statistics & Data Analysis*, Iowa University Press.
- CARLOS A. DE B. PEREIRA & SÉRGIO WECHSLER (1993), On the concept of P-value, *Brazilian J. Probability and Statistics*, v. 7, 159-77.
- Carlos A. de B. Pereira & Julio Stern (1999), Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy Journal* v. 1, 104-115.

Abstract

The focus of this paper is some fundamental concepts which are usually presented in a statistics course. Several examples are discussed to show how their use can produce incoherent conclusions. The examples illustrate the following interesting facts: independent and identically distributed random variables are strongly associated, conditional probabilities can depend on the specification process of the conditioning information, confidence intervals can indicate loss of precision when more information is added to the analysis and hypothesis test do not always reject unfeasible hypothesis.

Comentário a respeito de “Conceitos Estatísticos: Reflexões”, de Carlos Alberto de Bragança Pereira

*Marco Antonio Rosa Ferreira**

1. Introdução

O Professor Carlos Alberto de B. Pereira apresenta sua visão sobre importantes conceitos estatísticos que, em geral, são ensinados de forma imprecisa. O Professor Pereira discute com muita autoridade as diferenças entre vários conceitos: independência probabilística e estatística; intervalos de confiança e credibilidade; significância e evidência.

As idéias e os exemplos apresentados pelo Professor Pereira são muito claros, e apontam para pontos de divergência entre as duas principais correntes de pensamento estatístico: a clássica e a Bayesiana. A principal diferença conceitual entre estas duas correntes é relacionada aos condicionamentos. Enquanto cálculos da Estatística Clássica são realizados condicionalmente aos valores dos parâmetros, os cálculos da Estatística Bayesiana são feitos condicionalmente à amostra observada. Esta diferença permeia todos os exemplos e discussões do artigo do Professor Pereira. Claramente ele vê mais

* Endereço para correspondência: UFRJ – Instituto de Matemática – Caixa Postal: 68.530, Cep. 21945-970, Rio de Janeiro – RJ.

sentido na visão Bayesiana de cálculos condicionados no que de fato se conhece, ou seja, a amostra observada. Eu partilho desta visão do Professor Pereira.

Em minha opinião, a estatística caminha para um cenário em que os cálculos e interpretações de resultados serão predominantemente Bayesianos, mas em que vários conceitos Clássicos serão considerados, conceitos, tais como, erro quadrático médio freqüentista, cobertura freqüentista de intervalos de credibilidade, e consistência de estimadores. Outros conceitos freqüentistas perderão importância, como o conceito de estimador não viciado. Assim, eu vou aproveitar a oportunidade para discutir alguns destes conceitos.

Na seção 2, eu discuto porque estimadores não-viciados devem ser usados com cautela. Na seção 3, eu discuto alguns bons conceitos Clássicos e Bayesianos que acredito serão amplamente utilizados no futuro.

2. Falhas de estimadores não-viciados

Nesta seção, eu discuto duas situações em que estimadores não-viciados são inapropriados. A primeira é um exemplo fornecido por DeGroot (1986) referente à estimação de uma probabilidade. A segunda situação é em análise de regressão, onde a prática comum e salutar de eliminação de regressoras não significativas em geral implica na utilização de estimadores viciados.

Um exemplo interessante do perigo do estatístico somente utilizar estimadores não-viciados é apresentado no seguinte exemplo:

Exemplo (DeGroot 1986). Considere uma variável aleatória X com distribuição de Poisson com média λ . O interesse é a estimação de $\theta = e^{-2\lambda}$. Note que $0 < \theta < 1$ é a probabilidade de duas observações futuras estatisticamente independentes com distribuição de Poisson (λ) serem iguais a zero. É fácil mostrar que o único estimador não-viciado de θ dado X é $\delta(X) = (-1)^X$. Ou seja, este é um exemplo em que o único estimador não-viciado de uma probabilidade pode assumir somente dois valores, -1 ou 1 . Este é obviamente um péssimo estimador!

Uma solução Bayesiana é assumir *a priori* $\lambda \sim \text{Gama}(\alpha, \beta)$, $\alpha > 0, \beta > 0$, e utilizar a média *a posteriori* como estimador de θ . É fácil mostrar que *a posteriori* de λ é $\text{Gama}(\alpha + X, \beta + 1)$, e uma integral simples leva a

$$\hat{\theta} = E[e^{-2\lambda} | X] = \left[\frac{\beta + 1}{\beta + 3} \right]^{\alpha + X}.$$

Note que este estimador Bayesiano pertence ao intervalo [0;1], como é desejável para um estimador de uma probabilidade. Caso o analista não saiba como elicitar *prioris* ou queira utilizar uma *priori* não informativa, ele pode utilizar os valores $\alpha = \beta = 0$ o que são equivalentes *a priori* de Jeffreys para este problema. Neste último caso, $\hat{\theta} = 3^{-X}$.

Uma outra situação mais prática em que bons estimadores viciados são amplamente utilizados está em análise de regressão linear. Eu vou me referir aqui ao modelo verdadeiro como o modelo a partir do qual os dados foram gerados. Não vou aqui entrar na discussão filosófica sobre a existência ou não de modelos verdadeiros. Assim, em análise de regressão linear, o modelo verdadeiro seria o modelo que incluiria todas as variáveis regressoras que causam impacto na variável dependente e que não incluiria qualquer variável regressora que não causa impacto.

Os resultados sobre ausência de viés dos estimadores de mínimos quadrados para modelos com erros Gaussianos são válidos quando o modelo ajustado é o modelo verdadeiro. Do contrário, quando o modelo ajustado é diferente do verdadeiro, os estimadores de mínimos quadrados dos coeficientes das variáveis regressoras incluídas no modelo ajustado são, em geral, viciados (exceto sob condições de ortogonalidade das regressoras). Acontece que ninguém sabe de antemão qual é o modelo verdadeiro.

A situação mais comum na prática é a de haver muitas variáveis explicativas possíveis, e o que se faz usualmente é selecionar um modelo segundo algum critério de performance, por exemplo AIC, BIC, teste de significância, ou Full Bayesian Significance Test – FBST. Para detalhes sobre a utilização do FBST, ver Pereira e Stern (2000) e Pereira e Stern (2001). Além disso, o mais comum na prática é que as variáveis regressoras não sejam ortogonais. Assim, quando há muitas variáveis regressoras

candidatas e o número de observações não é muito alto, a probabilidade do modelo selecionado ao final de uma análise de regressão ser diferente do modelo verdadeiro é altíssima e, assim, os estimadores dos coeficientes das variáveis regressoras serão em geral viciados. Mas, então, por que todos nós eliminamos variáveis regressoras não significativas de nossos modelos? Nós estatísticos fazemos isto principalmente para que o modelo tenha boa capacidade preditiva, ou seja, para que o erro de predição de observações futuras seja pequeno. Assim, em nossa vida profissional estamos dispostos a admitir um pouco de viés em troca de muito mais precisão.

3. Alguns bons conceitos estatísticos

Nesta seção discutirei alguns conceitos estatísticos Clássicos e Bayesianos que me parecem muito importantes e que provavelmente serão amplamente utilizados no futuro.

Considero bons conceitos freqüentistas os conceitos de erro quadrático médio e cobertura freqüentista. Estes conceitos devem ser considerados quando o pesquisador está desenvolvendo um procedimento estatístico que será utilizado de forma automática em problemas práticos em diferentes contextos por pessoas que não têm conhecimento estatístico suficiente para elicitar distribuições *a priori*. Nestes casos, o procedimento deve ser bom para o conjunto de pessoas que o utilizarão. Portanto, o que importa são as propriedades do procedimento quando ele é utilizado muitas e muitas vezes, ou seja, as suas propriedades freqüentistas. Note que, como discuto no parágrafo a seguir, isto somente é verdade se a pessoa que utiliza o procedimento não aprende sobre o contexto em que ele é utilizado. É interessante notar que há vários exemplos em que estimadores Bayesianos baseados em *prioris* não informativas possuem melhores propriedades freqüentistas do que estimadores clássicos como o estimador de máxima verossimilhança (e.g., Ferreira e De Oliveira 2004, Fonseca, Ferreira e Migon 2005).

Penso que o mais interessante conceito Bayesiano seja o de informação *a priori*. A utilização de *prioris* permite que informação de especialistas e informação acumulada com experimentos anteriores sejam utilizadas para aumentar a precisão das análises estatísticas. Além disso, caso o analista freqüentemente analise dados em um mesmo

contexto, há procedimentos Bayesianos baseados em *prioris* subjetivas (Garthwaite et al. 2005), em modelos hierárquicos (Carlin e Louis 2000) e em modelos dinâmicos (West e Harrison 1997) que podem ser utilizados para se aprender sobre o contexto e, assim, incorporar na análise estatística a informação aprendida. Uma clara situação de vantagem de procedimentos Bayesianos é quando, em uma particular indústria, cálculos de risco devem ser repetidos periodicamente. Por exemplo, as seguradoras de automóveis podem aumentar em muito a precisão de seus cálculos de risco de sinistro através da incorporação da informação de anos anteriores mediante distribuições *a priori*. Isto pode trazer grandes vantagens competitivas para empresas que utilizem procedimentos Bayesianos.

Referências bibliográfias

- CARLIN, B. P. e LOUIS, T. A. (2000), *Bayes and Empirical Bayes Methods for Data Analysis*, CRC Press.
- DEGROOT, M. H. (1986), *Probability and Statistics*, Addison Wesley.
- FERREIRA, M. A. R. e DE OLIVEIRA, V. (2004), Bayesian analysis for a class of Gaussian Markov Random Fields, *Relatório Técnico 177*, DME – UFRJ.
- FONSECA, T. C. O, FERREIRA, M. A. R. e MIGON, H. S. (2005), Bayesian non-informative analysis for the Student-t regression model, *Relatório Técnico*, DME-UFRJ
- GARTHWAITE, P. H., KADANE, J. B. e O'HAGAN, A. (2005), Statistical methods for eliciting probability distributions, *Journal of the American Statistical Association* 100, pp 680-701.
- PEREIRA, C. A. B. e STERN, J. M. (2000), Evidence and credibility: full Bayesian significance test for precise hypotheses. *Entropy* 1, pp 69-80.
- PEREIRA, C. A. B. e STERN, J. M. (2001), Model Selection: full Bayesian approach, *Environmetrics* 12, pp 559-568.
- WEST, M. e HARRISON, J. (1997), *Bayesian Forecasting and Dynamic Models* (2nd Edition), Springer-Verlag: New York.

Discussão do artigo “Conceitos Estatísticos: Reflexões”, de Carlos Alberto de Bragança Pereira

*Rosangela Helena Loschi**

O professor Carlinhos, neste artigo, usa de toda a sua maestria para discutir de maneira simples e clara conceitos fundamentalmente importantes para a Estatística e que nem sempre são de fácil compreensão. Mais do que discutir o trabalho do professor Carlinhos, pretendo ressaltar aqui alguns pontos que são encontrados nas entrelinhas deste artigo e que julgo merecer um pouco mais de atenção.

A inferência Estatística na visão de muitos autores nada mais é do que uma metodologia coerente para se fazer previsões, condicionalmente nos dados observados. Ou seja, acredita-se que, mais do que saber a proporção de pessoas com uma certa enfermidade na população, o médico está mais interessado em avaliar a probabilidade de seu próximo cliente ser portador da enfermidade, $X_{n+1} = 1$, uma vez que vários de seus pacientes já a apresentaram. Sob independência, esta probabilidade é dada por:

$$P(X_{n+1} = 1 | X_1 = x_1, \dots, X_n = x_n) = P(X_{n+1} = 1),$$

* Endereço para correspondência: UFMG – Departamento de Estatística, Cep. 31270-901, Belo Horizonte – MG.

ou seja, é a mesma probabilidade que seria atribuída a $X_{n+1} = 1$ antes de observarmos qualquer paciente. Perceba que, sob independência, mesmo que os n pacientes já consultados pelo médico sofressem da enfermidade, esta informação não modificaria a expectativa dele sobre o que deverá ocorrer com seu próximo paciente. Por outro lado, se assumirmos a independência condicional, isto é, se consideramos que X_1, \dots, X_n, X_{n+1} são independentes condicionalmente numa variável aleatória Z que denota a proporção de pessoas com a enfermidade na população e se, além disto, assumirmos que nada se sabe sobre Z de tal forma que a incerteza sobre ela é descrita por uma distribuição uniforme no intervalo $(0, 1)$, tem-se que:

$$P(X_{n+1} = 1 \mid X_1 = x_1, \dots, X_n = x_n) = \frac{k+1}{n+2},$$

em que k denota o número de pacientes enfermos entre os n já observados, ou seja, neste caso, a informação sobre os pacientes já observados modifica a expectativa do médico sobre o que deve ocorrer com seu próximo paciente. Note que, ao assumirmos independência ao invés de independência condicional, estamos assumindo a impossibilidade de aprendermos com a experiência. Isto pode ser considerado razoável?

Outros conceitos importantes contrastados pelo professor Carlinhos são: confiança *versus* credibilidade e significância *versus* evidência. Além de toda a dificuldade envolvida na interpretação do conceito de confiança (que para mim é um dos conceitos menos intuitivos em Estatística), este artigo mostra vários exemplos em que o intervalo de confiança e o p -valor levam a conclusões diferentes do que aquelas indicadas pelo bom senso – amostras mais informativas deveriam gerar inferências melhores. Um dos grandes problemas que vejo com estes dois procedimentos clássicos de inferência é que ambos são construídos tomando-se como base o “princípio da amostragem repetida”, isto é, assume-se que o processo de amostragem pode ser repetido infinitas vezes sobre condições idênticas. Este princípio, por si só, é um pouco polêmico pelas condições que impõe, mas deixemos de lado esta discussão e pensemos sobre o p -valor, por exemplo, que é o que nos interessa neste artigo. Ao calcularmos o p -valor, percorremos parte do

espaço amostral uma vez que para sua obtenção calculamos a integral da distribuição da estatística de teste numa região deste espaço. Perceba que ao fazermos isto estamos levando em conta o mecanismo de geração dos dados e, automaticamente, valores para a estatística de teste que poderiam ter sido observados, mas não foram, e não só a informação trazida pela amostra. Este mecanismo de geração dos dados deve ser relevante para o procedimento de inferência (isto é, ele traz informação sobre o parâmetro de interesse?) ou o que deve realmente ser considerado é a informação trazida pela amostra observada? No cálculo da evidência da hipótese nula e, também, do intervalo de credibilidade, somente a informação da amostra de fato observada é que é considerada na análise.

Comentário sobre o artigo “Conceitos Estatísticos: Reflexões”, de Carlos Alberto de Bragança Pereira

*Gilberto A. Paula**

Em primeiro lugar gostaria de agradecer ao Prof. Francisco Louzada-Neto pelo convite para comentar o artigo do Prof. Carlos Alberto de Bragança Pereira submetido à *Revista Brasileira de Estatística - RBEs*. Vivemos tempos de grandes transformações que atingem toda a humanidade e com reflexos na formação dos nossos alunos que estão sendo educados num ambiente de muita informação e competição e que, mesmo não conhecendo os verdadeiros mecanismos geradores e motivadores de determinadas metodologias, conseguem assimilá-las de forma surpreendente. Todavia, como bem observado pelo Prof. Bragança Pereira, nota-se um certo desinteresse por parte desses alunos por disciplinas que envolvem conceitos lógicos e abstratos; o tempo disponível para estudos tem sido priorizado em disciplinas mais profissionalizantes em detrimento de outras que focam mais a fundamentação. Creio que esse comportamento seja em grande parte reflexo das exigências do mercado de trabalho por profissionais que conheçam uma grande variedade de técnicas e manipulem os principais aplicativos

* Endereço para correspondência: Univ. de São Paulo – USP, IME, Cep. 05315-970, São Paulo – SP.

estatísticos. Acrescento às carências apontadas pelo Prof. Bragança Pereira na formação dos nossos alunos a ausência de disciplinas de análise de dados, principalmente em nível de pós-graduação. Nota-se nos principais periódicos internacionais de Estatística uma exigência cada vez maior de exemplos interessantes, bem analisados e motivadores da metodologia desenvolvida. Assim, um pouco mais de fundamentação e análise de dados aliados ao forte enfoque metodológico que tem sido adotado poderá aprimorar a formação dos nossos alunos para enfrentarem os desafios que se apresentam nos dias atuais.

O Professor Bragança Pereira discute didaticamente vários exemplos mostrando a importância do uso das informações disponíveis para melhorar a qualidade de uma análise estatística. Em particular, no tópico *significância e evidência*, o Prof. Bragança Pereira apresenta um exemplo de teste de hipóteses em que as evidências contra a hipótese nula são muito distintas sob os enfoques clássico e Bayesiano. O referido professor utilizando informações *a priori* a respeito da hipótese alternativa obtém um *p-valor* modificado (denominado *pv*) que, diferentemente do *p-valor* obtido sob o enfoque freqüentista, aponta contra a hipótese nula. Creio que um olhar no poder do teste deste exemplo para um nível de significância de 5% poderia diminuir a discrepância entre as duas conclusões. Diversos outros exemplos interessantes envolvendo conceitos de *independência, informação, dependência condicional, confiança e credibilidade* são discutidos ao longo do texto.

Para finalizar, lembro que há 20 anos fiz uma disciplina de *estatística avançada* ministrada pelo Prof. Bragança Pereira que me pediu para apresentar um seminário sobre ancilaridade, conceito que me tem sido muito útil na disciplina de *modelos lineares generalizados*, por exemplo quando compara-se modelos binomiais condicionais com modelos de Poisson condicionais. Como é bem conhecido, se temos duas binomiais independentes $Y_1 \sim B(n_1, \pi_1)$ e $Y_2 \sim B(n_2, \pi_2)$ a razão de chances (*odds ratio*) definida por $\psi = \pi_1(1 - \pi_2) / \pi_2(1 - \pi_1)$ é uma medida de associação de grande interesse prático. Tradicionalmente aplicada na área médica, essa medida tem sido estendida para outras áreas do conhecimento, por exemplo finanças e seguros. A fim de se encontrar uma distribuição que dependa somente de ψ (crucial em estudos emparelhados), é usual

trabalhar com a distribuição condicional de Y_1 dado a estatística suficiente $Y_1 + Y_2 = m$ do parâmetro de incômodo, resultando numa hipergeométrica não-central. Porém, as inferências dos dois modelos (condicional e não-condicional) para ψ não são equivalentes. Por outro lado, se temos duas Poissons independentes $Y_1 \sim P(\lambda_1)$ e $Y_2 \sim P(\lambda_2)$ e, o interesse é fazer inferência para a razão de médias $\psi = \lambda_1 / \lambda_2$, podemos aplicar o mesmo procedimento para obter a distribuição condicional dado $Y_1 + Y_2 = m$ que é uma binomial; neste caso as inferências para ψ são equivalentes. Qual a diferença? No primeiro caso, a estatística suficiente não é ancilar para ψ , enquanto no segundo caso é (*vide*, por exemplo, Lehmann, 1994, pp. 546-547). Exemplos similares a este em que o conceito faz a diferença podem ser encontrados no desenvolvimento de outros procedimentos estatísticos e poderiam ser apresentados e discutidos mesmo em disciplinas mais profissionalizantes.

Referência complementar

LEHMANN, E. L. (1994). *Testing Statistical Hypotheses*. Second Edition, Chapman & Hall.

Réplica: Carlos Alberto de Bragança Pereira

O Professor Louzada-Neto me faz uma grande homenagem ao solicitar aos três ilustres professores comentários sobre meu artigo. Sinto-me realmente lisonjeado com o privilégio de ter colegas desse porte analisando algo que escrevo. Como é parte do processo a resposta aos comentaristas, apresento um complemento a nossa discussão. Agradeço aos três amigos, Rosangela, Gilberto e Marco, a honra que me proporcionam.

Rosangela complementa o artigo apresentando sua visão preditivista da estatística. Devo acrescentar que essa talvez seja adequada em muitos dos problemas do dia a dia do uso da estatística. No momento que estava escrevendo essa réplica me telefonou um dos ex-alunos com um problema interessante sobre riscos operacionais de uma instituição bancária. Usando a visão preditivista, justamente como a Rosangela apresentou, o problema se tornou trivial. Evidentemente, não poderia deixar de lembrar o cuidado que devemos ter com a precisão da linguagem. A Rosangela comenta sobre a informação contida na amostra. Seria interessante podermos sempre responder as seguintes perguntas: Informação sobre o quê? Onde está a informação? Quanto de informação? Finalmente, o que é informação? Uma discussão sobre esse conceito é a minha sugestão para uma próxima discussão.

Gilberto comenta sobre os fundamentos da estatística e que muitas vezes preocupa tanto a mim quanto a ele. Devo dizer que, embora não seja usuário da estatística Bayesiana, o professor Gilberto me fez um admirador de seus artigos e teses que orienta. Faz pouco tempo que precisei de um resultado importante e que não conseguia demonstrar. Fui direto pedir ajuda ao amigo. Ele me enviou parte da tese de

um de seus alunos que apresentava discussão perfeita sobre a aproximação de uma gama por uma log normal. Na verdade eu estava discutindo o modelo logístico normal para tabelas de contingência. Com isso, um de meus alunos entregou sua tese no último dia 14 de fevereiro. Minha convivência com Gilberto é de admiração. Admiração por ele ter uma das mais brilhantes produções acadêmicas, construída plenamente em nosso departamento, às vezes solitariamente, às vezes com seus alunos e outras em colaboração com seus colegas. Como chefe de departamento, posição que muitas vezes exerceu, soube harmonizar as divergências que sempre existem. Assim, aproveito aqui esse fórum para ressaltar que o Gilberto é o acadêmico ideal de qualquer universidade: excepcional professor, acadêmico e administrador.

Como todos já sabem Marco Antonio é um jovem pesquisador que já ocupa espaço de destaque em diversos grupos internacionais. Gostaria de lembrar que já recebeu um prêmio por seu trabalho na reunião da ASA de dois anos atrás. Na seqüência dedico-me a discutir seu comentário. Também ele complementou minha discussão com exemplos relevantes. Contudo não poderia esquecer de lembrar ao leitor minha preocupação com a precisão da linguagem. Evidentemente, Marco mostrou toda sua *expertise* com a operacionalidade da inferência Bayesiana. Mas vejo problemas de linguagem em dois pontos no seu texto.

No exemplo da distribuição de Poisson, creio que há um exagero no seu uso para críticas à estatística clássica ou freqüentista. Considere a notação $x | \lambda \sim P(\lambda)$, que indica que a observação x , especificado um valor de λ , segue um modelo Poisson com média λ . Isto quer dizer que, conhecida a média λ , para o cálculo das probabilidades sobre os possíveis valores de x , deve-se usar o modelo Poisson. No caso de $x = 0$, a probabilidade seria $\exp\{-\lambda\}$. O que está sendo estabelecido é que, conhecido o valor de λ , para o cálculo da probabilidade do evento $\{x = 0\}$, use o modelo Poisson. Na verdade não se estima probabilidade, pois essa é uma medida proveniente de nossos pensamentos. Não devemos permitir que estimem nossa mente! Veja que o parâmetro λ é uma quantidade que existe e é de interesse e desconhecida: pode assim ser estimada. Use a média, m , de uma amostra que foi observada e calcule a probabilidade da próxima observação ser zero. Isto é, $P\{x = 0 | \lambda = m\} = \exp\{-m\}$. Essa seria a

resposta apropriada de um estatístico clássico acostumado ao uso de estimadores não viciados. Não devemos bater em bonecos que inapropriadamente chamamos de estatísticos clássicos. Os bons estatísticos clássicos não cometem erros como o apresentado pelo Marco Antonio ou como aquele dos intervalos de confiança da distribuição uniforme que apresentei.

Para terminar, devo dizer que modelo verdadeiro é aquele que acerta a predição. Se sempre tivéssemos oportunidade de encontrá-lo, poderíamos ser confundidos com Ele. Isso não é um problema filosófico como o Marco colocou. É um problema de precisão de linguagem. Vejam que, inicialmente, uma classe de modelos é considerada. O que se deve procurar é o melhor modelo dentro dessa classe. Chamar esse modelo de verdadeiro é um vício da linguagem estatística. Nada de filosofia envolvida! Contudo existem casos construídos por simulação onde podemos falar em modelo verdadeiro. Mas fenômenos verdadeiros serem gerados por modelos matemáticos é algo que não cabe no meu entendimento do universo. Talvez seja meu despreparo teórico que me impede de aceitar esse tipo de conjectura.

Peço desculpas aos amigos por discutir aspectos dos seus comentários que gentilmente ofereceram ao meu artigo. Gostaria que não pensassem ser indelicado, mas, que isso é parte do jogo quando artigos são colocados para discussão. Espero no futuro poder participar mais de discussões como essa, incentivada pelo Professor Louzada-Neto.

Réplica: Marco Antonio Rosa Ferreira

O que o Professor Pereira chama de imprecisões de linguagem são na verdade diferentes pontos de vista, e, portanto, a discussão é mais filosófica do que estatística.

Com respeito ao uso do termo “modelo verdadeiro”, acho que todos concordam que os dados foram gerados de algum processo. É uma parte grande dos estatísticos vai concordar em chamar o processo gerador dos dados de modelo verdadeiro. Eu acredito que para a grande maioria dos problemas, tudo o que podemos fazer é tentar aprender sobre algum aspecto específico do processo gerador dos dados, e isto significa que geralmente a classe de modelos que utilizamos na análise de um problema não inclui o processo gerador dos dados. Mas esperamos que alguns dos modelos considerados seja suficientemente próximo do processo gerador dos dados para ser útil. Obviamente, este modelo útil não pode ser chamado de modelo verdadeiro. Quanto ao exemplo de estimação de uma probabilidade, há duas perspectivas diferentes, uma de estimação e outra de predição.

Eu acho que a grande maioria dos estatísticos concorda que há algumas probabilidades que podem ser estimadas, como por exemplo quando temos $Y \sim \text{Binomial}(n, \theta)$ e estimamos θ . Dito de outra forma, há sistemas físicos que possuem probabilidades como parâmetros, ou seja, há probabilidades que são propriedades físicas destes sistemas.

Acho que a discordância entre eu e o Professor Pereira se relaciona ao fato de eu mencionar probabilidades de observáveis, e de uma perspectiva de predição deve-se

calcular estas probabilidades a partir da distribuição preditiva dos observáveis. Na minha notação, dever-ser-ia calcular, então, $P(X_1 = 0, X_2 = 0 | X)$, onde X foi o valor já observado e X_1 e X_2 são observáveis futuros. É fácil calcular esta probabilidade no exemplo:

$$P[X_1 = 0, X_2 = 0 | X] = \int_0^{\infty} P[X_1 = 0, X_2 = 0 | \lambda] f(\lambda | X) d\lambda ,$$

onde $f(\lambda | X)$ é a densidade *a posteriori* de λ .

Agora, note que se λ pode ser estimado, pois tem algum significado físico, então faz todo sentido admitir que qualquer função biunívoca de λ também pode ser estimada. Podemos pensar nesta função como uma reparametrização do modelo.

Acontece que $\theta = e^{-2\lambda}$ é uma função biunívoca de λ , e, portanto, faz sentido estimar θ . No meu comentário eu uso a média *a posteriori*:

$$\begin{aligned} E[\theta | X] &= \int_0^{\infty} \theta(\lambda) f(\lambda | X) d\lambda \\ &= \int_0^{\infty} \theta = e^{-2\lambda} f(\lambda | X) d\lambda \\ &= \int_0^{\infty} P[X_1 = 0, X_2 = 0 | \lambda] f(\lambda | X) d\lambda \\ &= P[X_1 = 0, X_2 = 0 | X]. \end{aligned}$$

Ou seja, a média *a posteriori* de θ é igual a probabilidade preditiva do evento $\{X_1 = 0, X_2 = 0\}$. Note que em nenhum momento eu estimei a probabilidade preditiva.

É assim que os dois pontos de vista, de predição e de estimação, estão relacionados. Não há imprecisão alguma na forma como eu coloquei o exemplo. Sei que é uma forma não usual para quem gosta de pensar em predições, mas é uma forma que em meu entendimento é perfeitamente válida.

Tréplica: Carlos Alberto de Bragança Pereira

Interessante é perceber que, subitamente, ficamos envolvidos em uma discussão fora do foco do artigo original, tema principal deste fórum, criado pelo editor da revista. Serei, então, breve nesta tréplica.

O Professor Marco usou grande parte de seu texto ensinando como se faz inferência Bayesiana, apropriadamente. Isso é bom, pois mostra aos leitores a forma como, Bayesianos brasileiros, trabalhamos desde os idos de 1970, quando iniciei meu caminho para a Bayesianidade. Assim, não vejo vantagem em discutir predição *versus* estimação. Filosofia não é parte, até o momento, de nossos interesses científicos: nem do Marco nem dos meus. A presente discussão foi motivada pela tentativa de complementação do artigo inicial, tendo recebido alguma atenção em minha réplica. A seguir completo os comentários de minha réplica em resposta à réplica do Marco.

Confirmando que meu interesse no artigo original foi criticar a imprecisão de linguagem dos professores de estatística. Por outro lado, creio que o Professor Marco também pecou com a precisão da linguagem. Usar uma estimativa como sua probabilidade não quer dizer que estimou *sua* probabilidade. Estimou, sim, uma função paramétrica e decidiu, por alguma razão, usá-la como sua probabilidade. Feliz ele que mostrou ser, naquele caso especial, coerente! Isto é, ao calcular corretamente *sua* probabilidade, obteve o mesmo valor nas duas formas de cálculo que apresentou. Isto não quer dizer que esse comportamento deva ser seguido em todas as ocasiões de acesso às probabilidades.

Imagine, o leitor, que eu e o Marco estávamos interessados nos valores de X_1 e X_2 , os números de células defeituosas que seriam obtidas em duas amostras de células

de dois ratinhos, de uma amostra de 20 ratinhos, que foram de nós escondidos. Note que, célula defeituosa é um evento raro em um animal vivo, caso contrário estaria morto. Em experimentos desse tipo, coletam-se em torno de 2 000 células para encontrar-se cerca de três ou quatro, em animais saudáveis. Assim assumir um modelo Poisson é bastante adequado. Dessa forma, o Marco usou o seu método com a amostra de 18 observados e estimou a função paramétrica $\theta = e^{-2\lambda}$. De fato sua estimativa foi 0,55, que usou como sua probabilidade *posteriori* para o evento $\{(X_1, X_2)\} = (0,0)$. Sua *estimativa* da probabilidade foi um pouco diferente do biólogo que pegou a média da amostra, 1/2, e fez um *plug-in* na função, encontrando 0,61. Achei tudo muito justo!

Entretanto, as pessoas não sabiam que o técnico do laboratório era meu conhecido e que, tendo ido a sua casa, percebi que ele havia levado os dois ratinhos para sua filhinha. Pedi, então, que ele realizasse a análise e me informasse o resultado, o que me custou 100 reais. O resultado foi $\{(X_1, X_2) = (1, 0)\}$. Assim minha probabilidade para o evento anterior, (0, 0), é zero e um para o ponto (1,0). Seria correto afirmar que, com 100 reais, consegui uma estimativa da probabilidade (verdadeira) superior à do Marco? Acho que a precisão da linguagem é importante: o Marco continua com uma boa estimativa do parâmetro θ , mas com uma péssima *estimativa* da probabilidade do ponto (0, 0).

Insisto que o problema discutido aqui não é o objetivo do artigo original. O problema que criei ao comentar o texto do Professor Marco foi porque afirmei que o exemplo da Poisson, por mais de décadas conhecido, não serviu para que os estatísticos frequentistas abandonassem o uso dos estimadores não-viciados. Por que não? Porque o estatístico não deve estimar modelos estatísticos! Deve sim estimar parâmetros e usar a estimativa no modelo, se desejar.

Finalizo essa tréplica ressaltando o fato de o Marco ter usado o argumento “a grande maioria dos estatísticos concorda” para convencer o leitor. Alguém que segue este argumento certamente deveria aceitar estimadores não-viciados. Mesmo que um Einstein esteja dentro de um grande grupo que concorda com algo errado, não transforma esse algo em correto. Afirmando que não há alguma discordância entre mim e o Marco do ponto de vista operacional. Tudo que ele mostrou ser razoável é o que nossos

Bayesianos brasileiros vêm fazendo por muito tempo. Nossa discordância é apenas com relação à precisão da linguagem.

Identificação dos fatores determinantes do crescimento em crianças com baixa estatura: uma aplicação do Modelo de Cox estratificado combinado com um modelo para dados longitudinais

*Magda Carvalho Pires**
*Enrico Antonio Colosimo***

Resumo

O Hormônio do Crescimento é um importante agente do desenvolvimento humano e, quando sua deficiência é diagnosticada, são ministradas doses periódicas de acordo com um acompanhamento médico. Este estudo consiste na análise de dados obtidos através do acompanhamento clínico de 81 crianças deficientes em hormônio do crescimento. O objetivo é identificar os fatores associados ao processo de crescimento dessas crianças. Uma análise estatística padrão consistiria em usar o Modelo de Cox, em que o evento de interesse é a criança atingir a altura alvo. No entanto, a grande maioria das crianças não atingiu essa altura até o final do acompanhamento, o que implicaria em um excesso de censuras. Assim, a análise proposta consiste de duas etapas: (1) ajustar um modelo para descrever o processo de crescimento e a partir dele prever o tempo até o evento e em seguida (2) utilizar o Modelo de Regressão de Cox. A análise final através do Modelo de Regressão de Cox estratificado por renda demonstrou que as variáveis raça, altura inicial, tipo de recém-nascido (AIG, FIG, GIG) e parto traumático (sim ou não) influem no tempo até a criança atingir a altura alvo. Além destas, a interação entre parto traumático e recém-nascido foi significativa.

Palavras-chave: Hormônio do crescimento, modelo não-linear, riscos proporcionais.

* Endereço para correspondência: Departamento de Estatística – UFMG - 31270-901, Belo Horizonte – MG, e-mail: magdacpires@yahoo.com.br.

** E-mail: enricoc@est.ufmg.br.

1. Introdução

O crescimento pode ser definido como o processo pelo qual indivíduos mudam em tamanho e forma em um período de tempo. O principal responsável pelo crescimento é o Hormônio do Crescimento Humano (hGH, GH ou Somatotropina), secretado pela glândula hipófise (Knoepfelmacher, 2002).

O crescimento dos indivíduos ocorre basicamente em seis fases, cada uma delas com uma velocidade distinta. Na fase intra-uterina, o crescimento pode ser afetado por fatores materno-placentários, genéticos e hormonais. Os fatores genéticos e hormonais persistem na infância (do nascimento à puberdade), junto a outros como nutrição, afeto, etc. que surgem na fase da adolescência (Knoepfelmacher, 2002).

As crianças com déficit acentuado na secreção/ação do GH possuem, em geral, uma defasagem estatural significativa, características crânio faciais peculiares (fronte olímpica, hipoplasia de face e nariz em sela), voz infantil, pele e cabelos finos, aumento da relação massa gorda/massa magra, hipoglicemina, redução da densidade óssea e atraso do início da puberdade (Knoepfelmacher, 2002).

A deficiência de Hormônio de Crescimento pode se manifestar em graus variados e ter muitas causas diferentes. Segundo Knoepfelmacher (2002), por ocasião do diagnóstico, faz-se uma avaliação da baixa estatura, buscando informações sobre a história do paciente, condições da gestação, parto e nascimento, alimentação, prática de atividade física, determinação da altura alvo, avaliação do desenvolvimento puberal e da velocidade de crescimento, além de um exame físico detalhado.

Em crianças com defasagem no crescimento, é feita uma avaliação para verificar se a baixa estatura é devida à má secreção/ação do hormônio do crescimento. Em caso positivo, são ministradas doses de hormônio sintetizado e o desenvolvimento de cada indivíduo é acompanhado a cada três meses.

Este estudo utiliza técnicas estatísticas para identificar fatores determinantes do crescimento de crianças com deficiência do GH. A análise consistiu de duas etapas: (1) um modelo estatístico é inicialmente ajustado para obter os tempos até as crianças atingirem a altura alvo e em seguida; (2) o modelo de riscos proporcionais de Cox (Cox, 1972) é utilizado para identificar os fatores determinantes do crescimento.

O artigo foi organizado da forma seguinte. A seção 2 descreve as informações registradas nas visitas ao consultório médico assim como a metodologia estatística. Na seção 3, ajusta-se um modelo longitudinal e estima-se o tempo necessário para que cada criança atinja a altura alvo. O ajuste do Modelo de Cox é realizado na seção 4. Discussões finais são apresentadas na seção 5.

2. Banco de dados e métodos estatísticos

O estudo foi realizado com 81 crianças participantes do Programa Hormonal do Crescimento da Secretaria de Saúde de Minas Gerais, diagnosticadas com deficiência do hormônio de crescimento, tendo um período de acompanhamento de no mínimo 19 meses em 31 de dezembro de 2002.

Na primeira visita ao consultório do médico responsável, foram coletadas informações como condições referentes ao parto, ganho de altura e características socioeconômicas da criança. O tratamento consistia na administração de hormônio do crescimento e o acompanhamento foi feito a cada três meses. A cada visita ao consultório, foram obtidas informações a respeito do desenvolvimento da criança (peso, altura, idade óssea e informações referentes à dosagem e aos efeitos colaterais do hormônio do crescimento).

No estudo foram coletadas 18 variáveis que foram consideradas potencialmente importantes para descrever o crescimento de crianças. As variáveis e seus códigos identificadores estão apresentados na Tabela 1. Uma variável que norteia a decisão de alta clínica é a altura alvo. Ela é definida pelos endocrinologistas pediatras como a média da altura dos pais subtraída de 7 cm para meninas e somada de 7 cm para meninos.

Tabela 1- Descrição das variáveis utilizadas no estudo

Variável	Categorias
Sexo	Masculino / Feminino
Raça (RACA)	Branca / Negra
Naturalidade	Grande BH / Interior / Outro estado
Tipo de parto	Normal / Cesário / Fórceps
Parto traumático (TRAUMA)	Sim / Não
Recém-nascido (RECEM)	AIG (adequado para a idade gestacional) / PIG (pequeno para a idade gestacional) / GIG (grande para a idade gestacional)
Apresentação	Cefálica / Pélvica
Renda	Até 2 SM / De 2 a 5 SM / de 5 a 10 SM / Mais de 10 SM
Diagnóstico / origem	Idiopático / Orgânico
Grau de deficiência	Isolado / DMHH (Deficiência múltipla de hormônios hipofisários)
Altura alvo	Entre 146 e 178 cm
Peso ao nascimento	Entre 1 250 e 4 240 g
Velocidade	Entre 0,5 e 5 cm/ano
Idade óssea	Entre 0,3 e 13 anos
Idade cronológica / inicial	Entre 2 e 21 anos
Altura inicial (IALtura)	Entre 71 e 154,8 cm
Peso inicial	Entre 7 500 e 58 000 g
Dose inicial	Entre 0,20 e 0,71 mml/kg

Como a grande maioria das crianças (77) não atingiu a altura alvo proposta no início do tratamento, muitas censuras seriam geradas ao utilizar técnicas de Análise de Sobrevivência que, por sua vez, acarretaria em uma análise pouco precisa.

Uma proposta alternativa é utilizar as várias medidas repetidas de altura de cada criança ao longo do estudo para prever o tempo necessário para que essa criança atinja a altura alvo. A partir dessas previsões, pode-se utilizar o Modelo de Cox para as observações reais e previstas conjuntamente.

Esta proposta é motivada pelos testes de degradação realizados em situações de engenharia (Meeker, Escobar, 1998). Nos testes de degradação, uma amostra de unidades industriais é estudada sob condições experimentais. A resposta de interesse medida não é mais o tempo até a falha, mas uma característica de performance do

produto. Estas medidas são registradas continuamente durante o período de teste, de forma que, ao final do teste, cada unidade tem a sua própria trajetória de degradação observada. A análise destes dados de degradação permite obter a estimação de quantidades de interesse. A principal vantagem dos testes de degradação sobre os testes clássicos baseados em tempos de falha é que a análise pode ser realizada muito mais cedo, mesmo sem um único tempo de falha. Ou seja, o grande benefício é a redução da duração dos testes de confiabilidade. Exemplos de uso destes testes podem ser encontrados em Yacout, Salvatore e Orecchwa (1996) e Oliveira e Colosimo (2004).

3. Modelo longitudinal

Como análise preliminar, é relevante explicarmos por meio de modelos matemáticos o desenvolvimento das crianças (curvas de crescimento). O ajuste dessas curvas para dados longitudinais consiste em descrever e sumarizar o processo de crescimento com um número limitado de parâmetros. Através desse ajuste, pode-se caracterizar o padrão de crescimento, tendo o mesmo significado para todos os indivíduos e, permitindo, portanto, a comparação entre os mesmos.

Haustic (1989) propõe alguns modelos para descrever o comportamento de curvas de crescimento. Neste estudo, foram selecionados quatro desses modelos: linear, polinomial de ordem 2, contagem e logístico. A expressão destes modelos está apresentada na Tabela 2.

Os três primeiros modelos são lineares nos parâmetros, sendo ajustados por mínimos quadrados no software estatístico SPSS. O último modelo é não linear nos parâmetros, necessitando de rotinas especiais para ser ajustado (foi utilizada a rotina LME do software S-PLUS – Pinheiro e Bates, 2000). As estimativas dos parâmetros foram obtidas considerando as alturas médias em cada visita. Para identificar o modelo que melhor se ajusta aos dados em questão, foram utilizados o critério numérico AIC de Akaike e a comparação gráfica dos valores observados *versus* preditos.

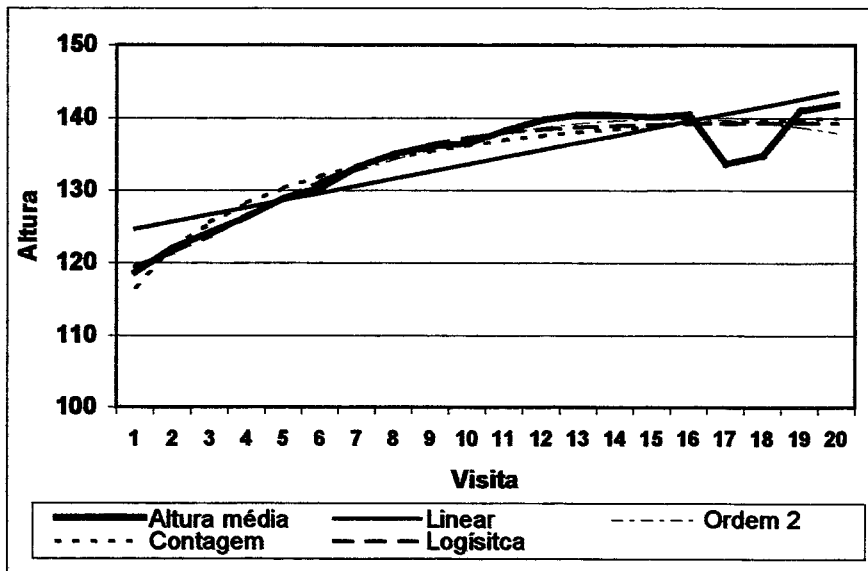
De acordo com o critério AIC (Tabela 2) e o gráfico dos valores preditos *versus* observados (Figura 1), com exceção do modelo linear, os outros três modelos explicam de forma adequada a altura média dos pacientes em função do tempo. Os modelos são

necessários para extrapolar o tempo no qual a criança atingiria a altura alvo. Este ponto requer considerações clínicas referentes ao comportamento esperado desta curva de crescimento além da trajetória observada. Ou seja, até em torno da vigésima visita (aproximadamente 60 meses) pode-se visualizar pelos dados o comportamento da curva de crescimento e, a partir deste valor, informações clínicas devem ser consideradas. Após uma análise da Figura 1 e Tabela 2, optou-se pelo modelo de contagem.

Tabela 2 - Comparação dos modelos ajustados quanto ao número de parâmetros e ao AIC

Modelo	Forma Funcional	Nº parâmetros (ρ)	AIC
Linear	$Y = \beta_0 + (\beta_1 * tempo)$	2	70,9%
Polinomial de ordem 2	$Y = \beta_0 + (\beta_1 * tempo) + (\beta_2 * tempo^2)$	3	89,9%
Contagem	$Y = \beta_0 + (\beta_1 * tempo) + \beta_2 * \ln(tempo+1)$	3	88,2%
Logístico	$Y = \beta_1 + \frac{\beta_2 - \beta_1}{1 + \exp[(\beta_3 - tempo) / \beta_4]}$	4	95,6%

Figura 1 - Comparação gráfica dos modelos quanto aos valores observados versus predito



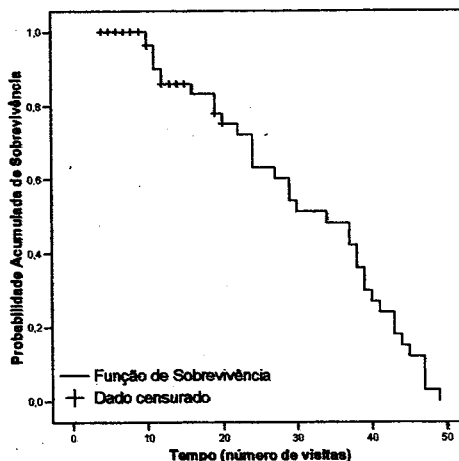
Das 81 crianças participantes, 32 delas haviam interrompido o tratamento, pois as avaliações médicas demonstraram que elas não atingiriam a altura alvo por razões diversas. Apenas 4 crianças atingiram a altura alvo durante o tratamento, antes do fim do estudo.

Para as demais 45 crianças, estimou-se o tempo em que elas atingiriam a altura alvo através da extrapolação do modelo estimado de contagem. Essa extrapolação foi satisfatória para 32 casos, mas em 7 deles não houve convergência do processo iterativo e em outros 6 casos o tempo estimado foi grande o suficiente para se considerar que clinicamente essas crianças não iriam atingir a altura desejada. A dificuldade em obter as estimativas para estas 13 crianças pode ser explicada por uma possível inadequação do modelo para elas ou mesmo situações em que estas crianças realmente não iriam atingir a altura alvo.

4. Ajuste do modelo de Cox

Através do modelo longitudinal, obteve-se 36 tempos de falha (4 realmente observados e 32 estimados pela curva de crescimento) e 45 observações censuradas. A função de sobrevivência estimada de Kaplan-Meier (Figura 2) forneceu um tempo mediano de 31 visitas (ou 93 meses) até que a criança atingisse a altura alvo.

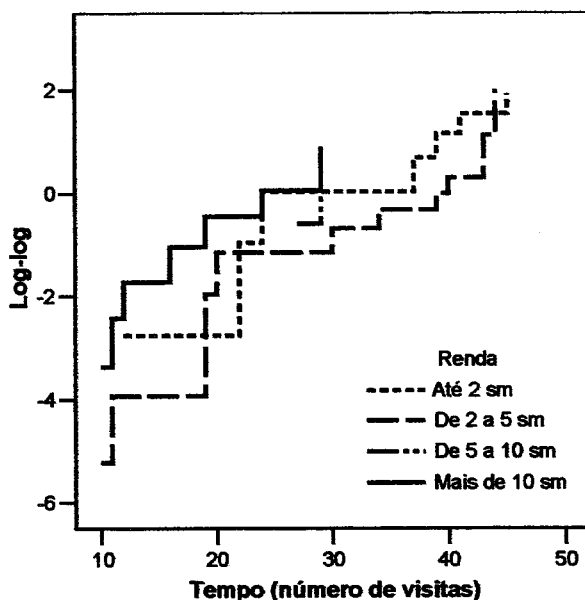
Figura 2 - Função de sobrevivência



Para selecionar quais das variáveis realmente fariam parte do estudo, utilizou-se uma estratégia de construção de modelos derivada daquela proposta por Collett (1994) e também utilizada por Colosimo (2001). Após utilização do método, as variáveis incluídas na análise final foram raça, ocorrência de parto traumático, recém nascido, renda e altura inicial, com a presença das possíveis interações entre essas variáveis.

Quando o modelo foi ajustado, percebeu-se através de técnicas gráficas, logaritmo da função de risco acumulada *versus* o tempo ou gráfico log-log (Klein e Moeshberger, 1997), que a suposição básica de riscos proporcionais não foi satisfeita para a variável renda (Figura 3).

Figura 3 - Curva do logaritmo da função de risco acumulada *versus* tempo para a covariável renda



Para solucionar esse problema, Kalbfleisch e Prentice (1980) propõem ajustar um modelo de Cox estratificado pela covariável que não possui riscos proporcionais. O ajuste para o modelo estratificado pela covariável renda é apresentado na Tabela 3.

Tabela 3 - Ajuste do modelo de Cox estratificado por renda

	Estimativa	Erro padrão	Valor-p	RR	IC (RR,95%)	
RACA	3,086	0,953	0,001	21,898	3,385	141,682
IALtura	0,078	0,023	0,001	1,081	1,033	1,131
TRAUMA	-0,538	1,068	0,614	0,584	0,072	4,734
RECÉM (GIG)			0,011			
RECÉM (PIG)	-4,373	1,607	0,007	0,013	0,001	0,294
RECÉM (AIG)	-2,559	1,567	0,103	0,077	0,004	1,670
RECÉM (GIG)*TRAUMA			0,021			
RECÉM (PIG)*TRAUMA	3,087	1,341	0,021	21,915	1,583	303,416

As seguintes interpretações podem ser obtidas desse modelo final:

- O risco de uma criança branca atingir a altura alvo é 22 vezes o risco de uma criança da raça negra. Além disso, podemos afirmar com 95% de confiança que esse risco varia entre 3,4 e 141,7;
- Para cada aumento de uma unidade na altura inicial, o risco de atingir a altura alvo é aumentado em 8%;
- Para crianças que tiveram parto traumático e que eram AIG (adequada para idade gestacional), o risco de atingir a altura alvo é cerca de 13 vezes o risco de crianças que tiveram parto traumático e que eram GIG (grande para idade gestacional); e
- Para crianças que tiveram parto traumático e que eram PIG (pequena para idade gestacional), o risco de atingir a altura alvo é 22 vezes o risco de crianças que tiveram parto traumático e que eram GIG (grande para idade gestacional).

5. Discussão

Em estudos que têm como foco o crescimento de indivíduos, a identificação de fatores determinantes de tal processo é fundamental para a prática clínica. Uma análise estatística nesta situação necessita de dados de sobrevivência destas crianças.

Entretanto, uma condição que ocorre com frequência é ter um período curto de acompanhamento devido a custos ou perda de acompanhamento das crianças. Este fenômeno gera um excesso de censuras que compromete a análise destes dados. Uma solução alternativa envolve registrar uma medida relacionada com o evento de interesse.

Na situação descrita, o próprio acompanhamento das alturas registradas da criança ao longo do período de tratamento foi utilizado para prever o tempo até o evento que, neste caso, era atingir a altura alvo. Esta previsão foi realizada a partir de modelos que descrevem o crescimento das crianças.

Um modelo estatístico foi utilizado para descrever a curva de crescimento e, a partir desse modelo, prever o tempo em que a criança atinge a altura alvo. Esse tempo estimado é utilizado em técnicas usuais de Análise de Sobrevida como o tempo em que o evento em estudo ocorreu.

A precisão das estimativas dos tempos para atingir a altura alvo não foi levada em conta na segunda etapa da análise conforme alerta um dos revisores. Dessa forma, é importante observar que os erros padrão das estimativas na Tabela 3 devem estar ligeiramente subestimados. Este é um tópico de pesquisa que estamos desenvolvendo na situação sob estudo. A forma de tratamento proposta na literatura (Meeker e Escobar, 1998) não se aplica na situação desse artigo pois temos uma amostra heterogênea (presença de covariáveis).

A partir do banco de dados composto por respostas observadas e estimadas, utiliza-se as técnicas usuais de análise de sobrevivência para responder as perguntas de interesse. No problema proposto, utilizou-se do modelo de Cox para identificar os fatores determinantes do crescimento em crianças. O resultado final mostrou que através do Modelo de Regressão de Cox estratificado por renda, as variáveis raça, altura inicial, tipo de recém nascido (AIG, PIG, GIG) e parto traumático (sim ou não) influem no tempo até a criança atingir a altura alvo.

Referências bibliográficas

- COLLETT, A. (1994). *Modelling Survival Data in Medical Research*. London: Chapman and Hall.
- COLOSIMO, E. A. (2001). *Análise de Sobrevida Aplicada*. 9º SEAGRO. Piracicaba.
- COX, D. R. (1972). *Regression Models and Life Tables (with discussion)*. J. R. Statist. Soc. B, 34, 187-220.
- HAUSTIC, R. C. (1989). *Mathematical Models for the Study of Individual Growth Patterns*. Rev. Epidem et Santé Publ. 37, 461-476.
- KALBFLEISCH, J. D. e PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.

- KLEIN, J. P. e MOESCHBERGER, M. L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer.
- KNOEPFELMACHER, MIRTA. (2002). *Deficiência do Hormônio de Crescimento (GH)* - Universidade de São Paulo.
- MEEKER, W. O. e ESCOBAR, L. A. (1998). *Statistical Methods for Reliability Data*. Wiley: New York.
- OLIVEIRA, V. R. B. e COLOSIMO, E. A. (2004). Comparison of Methods to Estimate the Time-to-failure Distribution in Degradation Tests. *Quality and Reliability Engineering International*, 20, 363-373.
- PINHEIRO, J. C. BATES, D. M. (2000). *Mixed Effects Models in S and S-Plus*. Springer.
- YACOLT, A. M., SALVATORE, S. e ORECHWA, Y. (1996). Degradation analysis estimates of the time-to-failure distribution of irradiated fuel elements. *Nuclear Technology*, 113, 17-189.

Abstract

The growth hormone (GH) is an important agent of the human development. Its deficiency is now a days regulated by a regular GH administration following a medical prescription. This paper considers a clinical study in which 81 GH-deficient children received replacement of GH. The aim of the study was to identify important factors associated with the growth process. A standard statistical analysis would consider fitting the Cox model in which the event of interest is the target height. However the great majority of the children did not reach the end point by the end of the follow-up period. This fact generated a large number of censoring observations. Our approach to the problem consisted of two steps: (1) fit a mixed effect model for the growth process in order to estimate failure times and (2) use the Cox model in the enlarged data set. The stratified Cox analysis identified race, initial height, type of newborn and delivery condition as associated with the survival response. An interaction term between delivery condition and type of newborn was also significant in the final model.

Agradecimentos

Os autores agradecem ao editor e aos revisores pela leitura cuidadosa e pelos comentários e sugestões tão valiosos, que contribuíram para a significativa melhora desse artigo.

Bootstrap estimation of standard errors in seasonal analysis

*R. P. Mentz**
N.M. Jarma
C. I. Martínez

Abstract

For the X-12-ARIMA seasonal adjustment procedure, standard errors of year-ahead seasonal estimates are computed by means of the bootstrap method. Direct estimates arise when the original estimated irregular is resampled, AR estimates arise when, before resampling, the correlation structure of an AR model is estimated and incorporated to the bootstrap irregulars. The calculations are done with four series covering various situations. Results are reasonably stable for two of the series with good seasonal fits, and less stable for the other two where fits are less satisfactory. The AR estimates are in general larger than the direct ones, as expected. The bootstrap procedures are suitable to be used as an optional calculation in the seasonal analysis.

Key words: Seasonal analysis, X-12-ARIMA, year-ahead seasonal, standard errors, bootstrap, autoregressive model.

* Endereço para correspondência: Universidad Nacional de Tucumán and CONICET - Casilla de Correo 209, (4000)Tucumán - Argentina, Fax 54-381-4364105, e-mail: mentz@herrera.unt.edu.ar.

1. Introduction

We consider the estimation of the standard errors of certain estimators arising in seasonal analysis. Among seasonal adjustment procedures, of singular importance is the family of X-11 procedures, originated in the USA Bureau of the Census in the early 1960's. It is well known that this procedure is statistically complex, being based upon ideas of smoothing by means of moving average filters, of iteration of sets of operations to which a series is subjected, and other techniques. Pfeffermann (1994) developed formulas for the standard errors of some estimates, by assuming a linear approximation to the seasonal adjustment procedure (Wallis, 1974). His analysis can take into account the sampling error of the survey estimates when relevant. We shall not pursue this interesting line of work in our paper. The bootstrap procedure appears as a reasonable alternative to obtain numerical estimations of some of the desired standard errors.

We use the bootstrap procedure to obtain estimations of standard errors of so-called year-ahead seasonal estimates. These are provided as part of the procedure, and consist of 12 monthly constants for monthly data, and 4 quarterly constants for quarterly data. The importance of these estimates is clear, since they are projections to be used in short term forecasting of the seasonal series.

We designed our study in terms of 4 observed series of different lengths, two arising in economic areas and two in meteorological contexts. The economic series are quarterly Gross National Product (GNP) and monthly Index of Industrial Production (IIP), both of Argentina; these are fairly regular series for which the seasonal procedures provide good fits. The meteorological series are two of those considered in Mentz et al (1999), namely monthly rainfall series for two locations in the State of Tucuman, Argentina, Las Cejas and La Cocha; in relation to the two economic series, these have less satisfactory seasonal fits.

The analysis is presented in two parts. In section 2 the bootstrap procedure is applied directly to the estimated irregular components of the series, which are the residuals of the seasonal fits. In section 3 the nature of these irregular components is analyzed in closer detail, and an alternative approach is used in which an autocorrelation structure is dealt explicitly.

2. Direct estimation of standard errors

2.1. The X-12-ARIMA seasonal adjustment program

The X-12-ARIMA seasonal adjustment program (Findley et al 1998) is a recent variant of the family of X-11 programs introduced in 1965 (Shiskin et al 1967). The mention of ARIMA in the name is due to the fact that for each given series a set of seasonal ARIMA models is automatically considered, so that if one of them provides a good fit (and is the best fit among the models available in the program), then one year of extrapolated data generated by that model is added at the end or beginning and end of the series, previous to the operation of the seasonal adjustment.

The program assumes a multiplicative model, $X_t = C_t S_t I_t$, where X is the observable time series, C is a trend cycle component, S is a seasonal (intraannual) component, and I is an irregular component. For many series a multiplicative model available in the program is preferred; in the case of positive components it can be reduced to the additive case by taking logarithms.

A complex computational procedure is used to produce final estimates \hat{C} and \hat{S} . A basic device used in the program is a set of moving average filters, which are symmetric in the case of observations in the middle of the series, and become asymmetric for observations near the ends of the series. Wallis (1974) argued that the combined operation of these filters can be taken (approximately) as represented by a long moving average. The use of ARIMA models for extrapolation will increase the number of months at which symmetric filters can be used.

The program includes a series of options that the user can consider, unless he chooses to use the automatic selections. Besides allowing for additive or multiplicative models and monthly or quarterly data, the program includes a treatment of extreme observations, the possibility of accounting for trading day variation, iterations of the procedures to produce preliminary estimates of the trend-cycle and seasonal components, use of a set of summary (diagnostic) measures, and others.

We concentrate in the results provided for the program, for monthly data, of 12 estimates of the seasonal component that can be used for prediction, and that are labeled *one year ahead seasonal factors*. The objective of this paper is to compute standard errors of these estimates, via the bootstrap procedure.

2.2. The Bootstrap procedure

Given an observed time series x_1, \dots, x_T with monthly observations, where $T = 12n$ for simplicity, we assume that this series has been generated by the *unknown probability model* $P = P(C, S, F)$, where F is the distribution function of the irregular component I . We assume that the indicated components generate the series by means of the additive model

$$X_t = C_t + S_t + I_t, \quad t = 1, 2, \dots, T. \quad (1)$$

or by a multiplicative model postulating a product of components.

In our presentation of the bootstrap ideas we follow Efron and Tibshirani (1993), in particular their Chapter 8. The fact that it can be assumed that a model holds for a series which is to be processed by one of the X-11 programs has been frequently considered explicitly in the literature; see, in particular, Cleveland and Tiao (1976).

For the given observed time series, the X-12-ARIMA program produces (final) estimates \hat{C} (Table D12), \hat{S} (Table D10), \hat{I} (Table D13), and one year ahead seasonal factors (Table D10A), that we denote by \hat{S}_{aj} , $j = 1, 2, \dots, 12$. The estimated \hat{S}_{aj} are standardized so that their mean is 100 for the multiplicative and 0 for the additive model. For further use we note that we operate with centered irregulars, in either the additive or multiplicative situations.

Our objective is to use the available information on C , S and I to estimate the standard error of the \hat{S}_{aj} . Formally we consider an *estimated probability model* $\hat{P} = P(\hat{C}, \hat{S}, \hat{F})$ corresponding to P defined above. Here \hat{F} is a discrete distribution obtained from the estimated irregular terms: they define an empirical distribution

function having jumps of size $1/T$ at the estimated values $\hat{I}_1, \dots, \hat{I}_T$. In the context of this model we let

$$se_F(\hat{S}_{aj}), \quad j = 1, 2, \dots, 12, \quad (2)$$

denote the standard errors of the statistics \hat{S}_{aj} .

The estimates of the $se_F(\hat{S}_{aj})$ that we propose to compute are denoted by $s\hat{e}_B(\hat{S}_{aj})$ (or by $s\hat{e}_B$ when there is no risk of confusion), and are calculated by means of the following steps: (a) Let $\hat{I}_1^*, \dots, \hat{I}_B^*$ be B bootstrap samples of size T from \hat{F} . That is to say, each one is a sample of size T selected with replacement from among $\hat{I}_1, \dots, \hat{I}_T$; (b) With a given set \hat{I}_b^* , we associate a bootstrap time series

$$\hat{X}_{tb}^* = \hat{C}_t + \hat{S}_t + \hat{I}_{tb}^* \quad t = 1, \dots, T \quad \text{and} \quad b = 1, \dots, B, \quad (3)$$

where \hat{C} and \hat{S} are the estimates coming from processing the data x_1, \dots, x_T with the X-12-ARIMA program; (c) With each bootstrap series, we use the X-12-ARIMA program to process it and compute $\hat{S}_{aj}^*(b)$, the corresponding one year ahead seasonals; and (d) We then define the mean of the bootstrap values by

$$\hat{S}_{aj}^*(.) = \frac{1}{B} \sum_{b=1}^B \hat{S}_{aj}^*(b) \quad (4)$$

and we have that

$$s\hat{e}_B(\hat{S}_{aj}) = \left\{ \frac{1}{B-1} \sum_{b=1}^B [\hat{S}_{aj}^*(b) - \hat{S}_{aj}^*(.)]^2 \right\}^{1/2}, \quad j = 1, 2, \dots, 12 \quad (5)$$

are our bootstrap estimates of the standard errors of the \hat{S}_{aj} .

Efron and Tibshirani (1993, pages 57-58) also consider robust estimates of standard errors, defined by

$$s\tilde{e}_{B,\alpha}(\hat{S}_{aj}) = \frac{\hat{S}_{aj}^{*(\alpha)} - \hat{S}_{aj}^{*(1-\alpha)}}{2z^{(\alpha)}}, \quad j = 1, 2, \dots, 12, \quad (6)$$

where $\hat{S}^{+(\alpha)}$ is the 100 α % quantile of the bootstrap replications, and $z^{(\alpha)}$ is the 100 α % percentile of a standard normal distribution.

2.3. Series used in our study

Since our final results are numerical estimates of the desired standard errors, we paid attention to the design of our calculations, in the sense that a collection of series was used in an attempt to cover several different scenarios. Some of the dichotomies that we considered are economic *versus* other types of series, monthly *versus* quarterly observations, multiplicative *versus* additive models, satisfactory *versus* poor seasonal fits.

The series we analyze are all from Argentine sources: (1) Gross National Product - GNP - at market prices, quarterly observations, 1980-1996, multiplicative model; (2) Index of Industrial Production - IIP, monthly observations, 1980-1996, multiplicative model; (3) Rainfall in Las Cejas, Tucuman, monthly observations, 1975-1989, additive model; and (4) Rainfall in La Cocha, Tucuman, monthly observations, 1982-1997, additive model. The economic series GNP and IIP have very good seasonal fits in terms of X-11-ARIMA, considerable better than the rainfall series.

For further information about the series see the Appendix, where we also report on some of the main results obtained in processing these series with the X-11-ARIMA program.

2.4. Results of the bootstrap procedure

In this section we report on the results obtained in using the bootstrap procedure in the manner indicated in Section 2.2. We used $B = 200$ in all cases, a number that can be considered satisfactory for our problem (cf Efron and Tibshirani, op.cit.). For some of the examples we compared the results obtained with $B = 100$, 200 and 300, and noted a great deal of stability.

The present Section 2.4 contains the analysis of what we call the *direct method*, to be distinguished from the autoregressive (AR) procedure developed in Section 3; the two procedures are compared in Section 4. For each series processed by the direct method we present a box plot of the bootstrap replicates (Figures 1, 3, 6 and 8) and a table summarizing in numerical form the main results (Tables 1, 2, 4 and 5); to facilitate the comparison of the two indicated procedures, each one of them contains the corresponding results for the direct and AR procedures. For the direct method we also present histograms, to facilitate the analysis of the symmetry and approximate normality of the empirical distributions (Figures 2, 4, 7 and 9), and the results of a comparison with a very simple regression estimation of the seasonal components and their standard errors (Tables 3 and 6 and Figures 5 and 10).

2.4.1. Economic series

2.4.1.1. Gross National Product – GNP

Figure 1 contains the box plots of the 200 bootstrap replicates of each of the one year ahead seasonal factors for quarters I, II, III and IV. Since the model is multiplicative, the seasonal factors cluster around 1; however, the traditional form of the output in these seasonal programs is to write them as percentages. Hence, our figures are expressed in terms of percentages.

Each box extends from the first to the third quartile, and also shows the median of the replicates. The whiskers have lengths equal to the 1.5 times the difference between the quartiles, and the stars correspond to replicates falling outside these limits. The starred replicates fell within 3 times the quartile differences, and hence can be qualified as distant (but not very distant). As we will see in comparison with the other cases, this is a rather regular diagram.

The same information is presented in Figure 2 by means of histograms. The four histograms have the same scales and the same numbers of classes to facilitate the comparisons. In general the histograms show that the distributions of the bootstrap replicates of the one year ahead seasonal factors do not depart strongly from symmetry.

On the basis of the bootstrap estimates, Table 1 presents the analysis corresponding to formulas (4), (5) and (6) of Section 2.2. The first two lines of the table compare the \hat{S}_{α_j} coming from the X-12-ARIMA program with the averages $\hat{S}_{\alpha_j}^*(.)$ of the bootstrap replicates: we find a great deal of coincidence. The third line contains the bootstrap estimates of the standard errors $s\hat{e}_{200}$ defined in (5), and the next three lines contain, respectively, the robust estimates of the standard errors, $s\tilde{e}_{200,\alpha}$ calculated for $\alpha=0,84, 0.90$ and 0.95 (we take $z^{(0.84)} = 1$). The differences are small, confirming the idea that the bootstrap distributions are fairly regular, that the influence of *outliers* is small.

2.4.1.2. Index of Industrial Production – IIP

Figures 3 and 4, and Table 2 contain the information corresponding to the IIP, in correspondence with Figures 1 and 2 and Table 1 for the GNP. The box plots are quite regular, with essentially all bootstrap replicates lying within 1.5 times the difference between quartiles. The 12 histograms in Figure 4 do not contradict the idea of approximate normality of these bootstrap distributions. The first two lines of Table 2 show a very good agreement in terms of means, while the differences between the bootstrap estimates of standard errors and the three robust estimates are small.

2.4.1.3. Analysis of the results

1. The two economic time series that we studied, quarterly GNP and monthly IIP, are very well suited to be treated by the X-12-ARIMA program in its multiplicative form. In the Appendix we argue that, in terms of some of the main figures of merit commonly used, the signals coming from the program are all in the indicated direction.

2. The bootstrap procedure works very well, at least to be judged by the close agreement between the one year ahead seasonal factors produced by the X-12-ARIMA program and the mean of the corresponding bootstrap replicates (Lines 1 and 2 of

Tables 1 and 2, respectively). We recall that seasonal factors are expressed (arbitrarily) as percentages.

3. The bootstrap replicates show variability, but by considering the box plots (Figures 1 and 3, respectively) we find that no large *outliers* are present. This is also confirmed by the relation between the two types of estimates of standard errors considered below.

4. The distributions of the bootstrap replicates show reasonable similarities with normal distributions, as seen in the histograms contained in Figures 2 and 4, respectively.

5. Tables 1 and 2 exhibit the numerical values of the bootstrap estimates of the standard errors of the one year ahead seasonal factors. We recall that the seasonal factors are expressed as percentages, so that the standard errors are also multiplied by 100. They are computed according to formula (5), and then according to formula (6) for $\alpha = 0.84, 0.90$ and 0.95 , which are robust estimators. In agreement with the preceding comments, the differences among these various estimates are in general small.

6. In the case of the GNP series, referring to the estimates coming from formula (5), we find that there are differences among the quarters: the smallest standard error is for the first quarter, the fourth quarter is larger, and the second and third quarters have similar standard errors and they are numerically the largest.

Bootstrap estimates of standard errors tend to move in consonance with the size of the corresponding statistic (or of the average over *bootstrap* replications), without being proportional. In fact they represent 0.41%, 0.46%, 0.47% and 0.44%, respectively.

7. In the case of the IIP series, referring to the estimates coming from formula (5), we find a good deal of coincidence among the estimates of standard errors: 7 of the 12 estimates can be rounded to 1.2.

8. While comparison of the \hat{S}_{aj} with the averages $\hat{S}_{aj}^*(.)$ is useful to evaluate the performance of the procedure, we do not have a similar benchmark to compare the numerical values of the bootstrap estimates of the standard errors. In view of the very regular structure of the seasonal patterns in our series, we performed the following calculations. To the series adjusted for trend (i.e. the original series divided into the final estimate of the trend cycle component), we fitted regression models without constant, having linear trend terms and dummy seasonal variables. The fits of these models are compared with the X-12-ARIMA estimates in Figure 5, and we see that the approximations are quite reasonable. The vertical scales in Figure 5 are not in percentages.

The advantage of these approximations is that for the regression models we have estimates of the standard errors obtained as part of the inferential part of the regression procedure. The numerical values of the one year ahead seasonal factors and their standard errors are compared in Table 3. The estimates of standard errors coming from the regressions are quite similar among themselves, which is a consequence of the approach (The seasonal dummies are orthogonal). We find that the agreement in the case of GNP is quite good, and better than for the IIP. What is important from our point of view is that the orders of magnitude of the standard errors give reassurance to the bootstrap estimates.

9. We conclude that for the two economic series that we considered, one with quarterly and one with monthly observations, which are acknowledged to be very suitable for the seasonal treatment programmed in X-12-ARIMA under multiplicative models, the bootstrap procedure works very well to produce numerical estimates of the standard errors of the one year ahead seasonal factors which are obtained as a regular part of the output of the program.

2.4.2. Rainfall series

2.4.2.1. Las Cejas

Figure 6 contains the box plots of the 200 bootstrap replicates of each of the one year ahead seasonal factors for the 12 months. The model is now additive, and the seasonal factors are expressed as percentages and centered at 0. The variability and presence of *outliers* is more notorious than in the economic series. In particular we find a great deal of variability (including *outliers* in some cases) in January, March, August, and November. The corresponding histograms in Figure 7 do not show noticeable departures from (approximate) normality.

The main results of the bootstrap procedure are in Table 4. There is some level of agreement between the estimated seasonal factors \hat{S}_{α_j} and the averages of the bootstrap replicates $\hat{S}_{\alpha_j}^*(.)$. In comparing the bootstrap estimates of standard errors corresponding to formula (5) appearing in the third line of the table, with the robust estimates in the fourth, fifth and sixth lines, there are some differences, with the first estimates being larger than the robust versions, as expected. In fact, the differences are larger for most of the months mentioned in the preceding paragraph as exhibiting large variabilities.

2.4.2.2. La Cocha

Figure 8 contains the box plots of the 200 bootstrap replicates of each of the one year ahead seasonal factors for the 12 months. In comparison with the box plots for the two economic series and the rainfall series for Las Cejas, the present one shows considerably more variability, and a great deal of outlying values. Note in particular the small *outliers* in the otherwise rainy month of January, opposite to the large *outliers* in the otherwise dry month of July.

The indicated nature of the bootstrap replicates is also apparent in the histograms of Figure 9: their central parts do not contradict the idea of approximate normality, but the presence of very large or very small values is clear.

These remarks help interpret the bootstrap results in Table 5. There are differences between the original estimates of the seasonal factors and the averages of the bootstrap replicates (Lines 1 and 2 of the Table 5). What is more important for our analysis, there exist large differences between the usual bootstrap estimates of the standard errors computed according to (5), and the robust estimates computed according to (6). For the months of January and July mentioned above, the differences are the largest, the usual estimates are larger than 30, while the robust estimates are around 20. One consequence of this fact, is that the usual estimates range from 20.5 for May to 32.1 for January, while the robust estimates show less variability: They range from 18.5 for October to 24.8 for November when $\alpha = 0.95$, from 18.1 for May to 22.7 for February when $\alpha = 0.90$, and from 16.3 for May to 21.8 for April when $\alpha = 0.84$.

2.4.2.3. Analysis of the results

1. The two monthly rainfall series that we studied, Las Cejas and La Cocha (Province of Tucuman, Argentina) are not particularly well suited for treatment with the X-12-ARIMA program. They are processed under the additive model because they have zero values, corresponding to months without rain. In the Appendix we argue that among the figures of merit commonly used, some indicate that a reasonable fit can be expected, while others reject the quality of the fit.

2. The bootstrap procedure also finds difficulties. For example, there is disagreement between the one year ahead seasonal factors produced by the X-12-ARIMA program, and the mean of the corresponding bootstrap replicates (Lines 1 and 2 of Tables 4 and 5, respectively).

3. The bootstrap replicates show variability, including the presence of large *outliers* (box plots in Figures 6 and 8). This is particularly important in the case of La Cocha, and is also confirmed by the relation between the two types of estimates of standard errors considered below.

4. The distributions of the bootstrap replicates show reasonable similarities with normal distributions in their central portions, but the presence of *outliers* is notorious in the histograms contained in Figures 7 and 9.

5. Tables 4 and 5 exhibit the numerical values of the bootstrap estimates of the standard errors of the one year ahead seasonal factors. They are computed according to formula (5) in the third line, and to formula (6) with $\alpha = 0.84, 0.90$ and 0.95 in the fourth, fifth and sixth lines. The differences among these various estimates are large. It makes an important difference whether we adopt the usual or any of the robust estimates of the standard errors, in particular in the case of La Cocha.

In a similar situation, Efron and Tibshirani (1993, page 69) comment: " The large values of $s\hat{e}_{200}$ for [some months] are seen to be caused by a few extreme values of [the bootstrap replicates]. The approximate confidence interval $\theta \in \hat{\theta} \pm Z^{(1-\alpha)}s\hat{e}$ will be more accurate with $s\hat{e}$ equaling $s\tilde{e}_{200,\alpha}$ rather than $s\hat{e}_{200}$, at least for moderate values of α like .843."

6. Taking the values in the fourth line of Tables 4 and 5 as the estimates of the standard errors of the one year ahead seasonal factors, we find that they range from 9.8 in April to 13.5 in February for Las Cejas, and from 16.3 in May to 21.8 in April for La Cocha.

7. The regression approximation proposed for the economic series can also be tried here (the model has now coefficients restricted to sum to 0). The results are in Figure 10 and in Table 6. This information is still of some use, in terms of orders of magnitude.

8. We conclude that for the two monthly rainfall series that we considered, which are acknowledged to be hardly suitable for the seasonal treatment programmed by the X-12-ARIMA under additive models, the bootstrap procedure finds some difficulties, and the analysis points out to the convenience of using robust estimates of the standard

errors of the one year ahead seasonal factors which are obtained as a regular part of the output of the program.

3. Taking into account an autocorrelation structure

3.1. Description of the problem

In section 2 we considered the theoretical model (1) for the observed time series, and called I_t the irregular component. This is in agreement with the practice established since the origin of the X-11 procedures (Shiskin et al, 1967). One may wonder why this component is not called an error or residual component, as is done in several areas of statistics, for example in time series or in regression analysis. The idea is that the error terms in these areas arise as random shocks affecting the theoretical model that will be observable. As such, they are usually assumed to have a constant expected value (any non-constant part being ascribed to the other components of the model), they may also be assumed to have a constant variance, and, more important, to be independent or at least uncorrelated. These specifications are often summarized in the designation of white noise stochastic process, and as such appear in the definition of ARMA and ARIMA models, in many linear models including regression and experimental design models, and elsewhere. One use of this type of specification is in estimation theory, where they serve as guidelines for the definition or procedures.

The question is then whether we can, at least approximately, consider that the irregular component entering into the model for a time series to be processed with one of the X-11 procedures, is a white noise process. The answer to this question is negative: in building up the X-11 family of procedures, no assumptions of this kind were made about the irregular. The X-11 procedures center their attention in a careful and flexible estimation of the main components in (1), namely the trend-cycle and the seasonal components. It then follows that the irregular is expected to be free of trend-cycle and seasonal components. However, no assertion is made *a priori* about the irregular being a white noise process.

In view of this argument, we find that the approach used in section 2, that we called *direct*, will be adequate insofar as the irregular component is such that bootstrap samples from it retain the structure of the initial estimates. If this is not the case, we will find out that while \hat{I}_t is an autocorrelated time series, the bootstrap samples \hat{I}_t^* will be more nearly non-correlated, due to the effect of the random sampling used to define them. We will then say that the autocorrelation in the series of interest has not been preserved.

A proposal to deal with this possible source of difficulties was suggested by Findley et al (1990), and in essence consists in introducing in the bootstrap samples of irregulars, an autocorrelation structure estimated in the original estimated irregulars. This autocorrelation structure is that of an AR model fitted to the estimated irregular component, with an order estimated from the same data. In the bootstrap procedure developed upon this basis, it occurs that bootstrap samples are computed only of series that are expected to be white noise.

A different approach for this kind of problems, is to consider the so-called *moving block bootstrap* procedure, in which a length of the blocks is also to be determined empirically. See, for example, Efron and Tibshirani (1993), Li and Maddala (1996). We will not explore this possibility here.

3.2. Estimation of AR models and bootstrap procedure

As before (see Section 2.2), let \hat{I}_b be the estimated irregular component reported in Table D.13 of the output of the X-12-ARIMA program. We consider fitting to this series an invertible (causal) AR process of an order \hat{p} determined by the AICC criterion, which is Akaike's Information Criterion Corrected for bias (see, for example, Brockwell and Davis, 1991a, 1991b, that we follow in this section). The notation that we use is

$$\Phi_{\hat{p}}(\bar{B})\hat{I}_t = \hat{\varepsilon}_t, \quad (7)$$

where B is the backshift operator ($BZ_t = Z_{t-1}$),

$$\Phi_{\hat{p}}(\bar{B}) = 1 - \hat{\phi}_1 \bar{B} - \dots - \hat{\phi}_{\hat{p}} \bar{B}^{\hat{p}}, \quad (8)$$

and $\hat{\varepsilon}_t$ is white noise. We now invert (7) into

$$\hat{I}_t = \Phi_{\hat{p}}^{-1}(\bar{B}) \hat{\varepsilon}_t \cong \Pi_c(\bar{B}) \hat{\varepsilon}_t, \quad (9)$$

where

$$\Pi_c(\bar{B}) = 1 + \hat{\pi}_1 \bar{B} + \dots + \hat{\pi}_c \bar{B}^c, \quad (10)$$

is an approximation of finite order c to the infinite moving average defined by the inverse operator in (9). The order c is chosen so that $\hat{\pi}_s$ is small for $s > c$.

We now generate *bootstrap* samples of the $\hat{\varepsilon}_t$ that we designate by $\hat{\varepsilon}_b^*$, $b = 1, 2, \dots, B$. With these we generate the bootstrap irregulars,

$$\hat{I}_b^{**} = \Pi_c(\bar{B}) \hat{\varepsilon}_b^*, \quad t = 1, 2, \dots, T, \quad b = 1, 2, \dots, B, \quad (11)$$

and finally the bootstrap series

$$\hat{X}_b^{**} = \hat{C}_t + \hat{S}_t + \hat{I}_b^{**}, \quad t = 1, 2, \dots, T, \quad b = 1, 2, \dots, B, \quad (12)$$

We use two asterisks to distinguish from (3) of the direct method.

To the series in (12) we apply the X-12-ARIMA procedure to obtain the desired estimates of the year-ahead seasonals, and in turn their bootstrap estimate of standard errors.

We see that if the autocorrelation structure in the estimate irregular for a given series is important, the AR bootstrap series \hat{X}_t^{**} will differ from the \hat{X}_t^* considered in Section 2, and hence the estimates of the year-ahead seasonals may also differ, and also their estimated standard errors. The magnitude of this effect will be studied by comparing the results obtained by application of the two procedures.

3.3. Results of the empirical study

As indicated in Section 3.2, we start the empirical study by fitting to each time series an AR model of an order estimated from the same data. We used the programs in Brockwell and Davis (1991b), which make preliminar estimations of the parameters by using Burg's or Yule and Walker's procedures, then estimate the order by the AICC procedure, and finally produce the final estimates of the $AR(\hat{p})$ model parameters. We found the orders of 6 trimesters for GNP, 7 months for the IIP, 12 months for rainfall in Las Cejas, and 4 months for rainfall in La Cocha. We explored this last series in more detail, and found that similar results were obtained by using order 12 also for this rainfall series. In inverting the AR operator, we considered the way the coefficients decreased as the lag increased, and came out with the uniform suggestion of using lags up to $c=50$ in all 4 series. Hence, the irregular in (11) is generated by sums truncated at $c=50$, while we use three different values of the autoregressive order.

With these irregulars we generate the bootstrap time series defined in (12), obtain the estimates of the year-ahead seasonals, and their standard errors.

3.3.1 Bootstrap study of the economic series

The results of the bootstrap study for the quarterly GNP and monthly IIP are presented in tables and graphs with the same formats as those discussed in Section 2 for the direct method. The only exception is that we do not have tables like Tables 3 and 6, and figures like Figures 5 and 10, which contain regression analyses of the seasonal patterns.

Tables are 1 for GNP and 2 for IIP, and figures containing box plots are 1 for GNP and 3 for IIP. For the sake of simplicity we do not present the histograms corresponding to the bootstrap results under the AR procedure, since they lead to the same kind of conclusion that in the case of the direct method.

In view of the indicated circumstance we do not describe in detail these tables and figures, and postpone further discussion of these results until Section 4 in which they will be compared with those coming from the direct method.

3.3.2. Bootstrap study of the rainfall series

The numerical results of the bootstrap study appear in Tables 4 for Las Cejas and 5 for La Cocha, and box plots of the bootstrap replicates appear in Figures 6 for Las Cejas and 8 for La Cocha. The same comments of the preceding section apply to this case.

4. Comparing the direct and AR approaches

The comparison of the direct and AR approaches is facilitated by the approach followed in the presentation of the results in many of the tables and figures: in Tables 1, 2, 4 and 5 which summarize the main numerical results, and in Figures 1, 3, 6 and 8 containing the box plots of the bootstrap replicates, the results of the two approaches are presented in the same page.

In Tables 1, 2, 4 and 5 the 4 estimated standard errors are summarized by their medians and standard deviations: medians are used to compare positions, standard deviations to compare variabilities. Variability is related to the appearance of the corresponding box plot diagram, including the possible effect of *outliers*. Position and variability as measured in the tables, are then exhibited graphically in Figures 11 and 12.

GNP is the simplest series to analyze, partly because there are only 4 quarters to consider, and also because the series corresponds to a highly aggregated concept. Box plots for the AR(6) procedure show less variability than for the direct method (Figure 1), standard errors $s\hat{e}_{200}$ and medians of the 4 estimates of standard errors are higher for the AR(6) than for the direct method (Table 1), standard deviations of the estimates of se for the AR procedure are equal to or larger than those of the direct method in 3 of the 4 quarters. Hence, the graphical summary in Figure 11 can be interpreted as showing that, compared with the direct method, the AR(6) procedure produces larger estimates of $se(\hat{S}_{aj})$, its bootstrap replicates tend to exhibit less variability, and as a consequence its robust estimates $s\tilde{e}_{200,\alpha}$ differ less among themselves and from $s\hat{e}_{200}$.

For monthly IIP, $s\hat{e}_{200}$ under AR(7) is higher than under the direct approach in 11 of the 12 months (the exception is January), and medians are also higher in 11 months

(the exception is June). Hence, the conclusion is that AR estimates of standard errors tend to be higher than those of the direct procedure, which is confirmed by Figure 11. In terms of variability of the 4 estimates of $se(\hat{S}_{aj})$, and hence of the effect of the variability of the bootstrap replicates, standard deviations under AR are smaller than those of the direct method only in 6 of the 12 months.

As expected, the results for the rainfall series are less conclusive. We recall that these series have more variability than the economic ones, and that the X-12-ARIMA procedure found more difficulties in its fit to the data. For Las Cejas $s\hat{e}_{200}$ is larger under the AR(12) procedure in 8 of the 12 months, and the median of the 4 estimates of se is higher in 9 months, so that AR estimates of standard errors tend to be higher than those of the direct procedure, which is confirmed by Figure 12. Standard deviations of the 4 estimates of se are smaller for the AR procedure in 8 of the 12 months.

For monthly rainfall in La Cocha, $s\hat{e}_{200}$ is smaller under the AR(12) procedure in the 12 months, while the median of the 4 estimates of se is larger than for the direct approach only in 3 of the 12 months. This should be related to the shape of the box plot in Figure 8, and is also evident by observing Figure 12 and comparing La Cocha with the other 3 series considered in this study. Another consequence of these facts is that the standard deviations of the 4 estimates are smaller for the AR procedure in 9 of the 12 months.

5. Conclusions

This study deals with some estimation problems generated in applying the X-12-ARIMA procedure to time series of monthly or quarterly observations. The X-12-ARIMA procedure assumes that the series can be decomposed multiplicatively or additively into certain components; it also has available an option known as pseudo-additive. In the simplest case, these are the trend-cycle, seasonal and irregular components. A more detailed analysis of the given series may lead to consider other components: in a related resampling study, Findley et al (1990) considered in addition trading day, holiday and

outlier components, and treated them with a module of X-12-ARIMA called Reg-ARIMA (Findley et al 1989).

For any of these models, the program X-12-ARIMA produces estimates of the components and of other related quantities. Since one of the main objectives of seasonal analysis is forecasting, we consider the one year ahead seasonal factors estimated by the program. These are extrapolations of the seasonal component, which are recommended to be used in the analysis of the first year following the period of observation.

For these one year ahead seasonal factors, which consist of 12 estimates for monthly data and 4 estimates for quarterly data, we concentrate in estimating their corresponding standard errors. These are the key quantities needed to appraise the variability of the estimates. However, due to the complicated nature of the X-12-ARIMA procedure, no closed-form either exact or approximate expressions are available for these standard errors.

It then follows that the bootstrap procedure is *a priori* adequate to be used for standard error estimation. In this paper we study two procedures, namely: (a) the direct method, in which resampling is performed with the estimated irregular component estimated in the empirical decomposition, as is usually done in bootstrap studies dealing with (independent) cross-section data; (b) the $AR(\hat{\rho})$ procedure, in which the autocorrelation structure of the estimated residual is studied, and if it is found of importance, incorporated into the resampling procedure.

The differences between the two approaches are found to be important. The explicit consideration of the AR structure leads in general to more concentrated values of the bootstrap replicates, less need to consider robust estimates of the standard errors, and larger values of the estimated standard errors.

The procedures were applied to 4 different time series, two arising in the economic field (Argentine's quarterly GNP and monthly IIP), and two monthly rainfall series. The economic series are analyzed by means of multiplicative models, and the rainfall series by means of additive models because they have zero observed values in some months. The indications coming, when the X-12-ARIMA procedure is applied to these series, are that for the economic series the fit is quite good, while for the rainfall

series has considerably more difficulties. This can be regarded as a strength of our analysis, since it is possible to check the expectation that the bootstrap results will tend to be weaker for the rainfall series.

In view of the preceding considerations, in all cases the automatic (or default) options available in the X-12-ARIMA program are used. As indicated in the first paragraph of this section, a more detailed or specific model could be developed for each series. However, the rationale in the application of the bootstrap procedure will not change, since all estimated components will enter unchanged in the formation of the bootstrap replicates of the series, only the treatment of the irregular being subject to analysis.

Two main remarks can be offered in conclusion. One is concerned with the numerical value of the estimated standard errors. In terms of order of magnitude, we can use the summary measures available in Tables 1, 2, 4 and 5. Two possible selections are the medians (over periods) of the $s\hat{e}_{200}$ estimates, and the medians (over periods) of the four estimates of se . These results are as follows:

Box 1 – Medians of the four estimates of se and $s\hat{e}_{200}$, AR procedure

Series	Number of observations	Estimated AR order	Median of $s\hat{e}_{200}$, AR procedure	Median of medians of estimates of se , AR procedure
GNP	68	6	0.518	0.508
IIP	204	7	1.260	1.196
Las Cejas	180	12	13.786	13.073
La Cocha	192	12	21.173	19.644

As for overall consistency, these median values can be compared with the simple regression estimates in Tables 3 and 6, the values above being slightly larger.

The second remark is that a bootstrap procedure similar to that developed here, can be incorporated in the empirical application of programs like X-12-ARIMA, for the purpose of obtaining numerical estimates of the standard errors of interest. In this way they will complement the list of numerical estimates produced by this kind of programs.

APPENDIX. More information about the series used in this study

In Section 2.3 we introduced the four series studied in this paper. The two economic series (quarterly GNP and monthly IIP) are two standard measures used in economic description and analysis. The two monthly rainfall series are included in a study emphasizing cyclical analysis (Mentz et al, 1999). Brief descriptions follow.

GNP. It is a frequently-used indicator of the global behavior of Argentine's economy, computed and published quarterly. It is a weighted aggregate of economic series representing added value in the production of goods and services. It is often defined as the value (at market prices) of the production done in the economic space of a country, with the participation of residing and non-residing factors of production.

IIP. This weighted index is based upon series of the manufacturing sector of the economy, is produced by a private institution, Fundación FIEL, and comprises 61 products; often 100 is regarded as a desirable number of products. The design is of a longitudinal or panel type. Most of the information is gathered from associations and chambers of producers, which simplifies the procedure and reduces the costs. The selection of products is done according to two basic ideas: (a) that the available series provide good coverages for the selected products; (b) that the series are readily and regularly available.

Rainfall Series. These series are formed by monthly totals of rainfall in the indicated places. Those for Las Cejas were provided by a state-supported agricultural experimental station, while those for La Cocha were collected by a private farmer.

In our applications of the X-12-ARIMA procedure, in general the default or automatic options were used, when available; see our comments in Section 5. All series were processed in their original available scales, without introduction transformations; the possibility of transforming the rainfall series was considered in Mentz et al (1999), and found that they did not imply improvements in the application of the procedures.

Appendix Table 1 contains information about the results of applying the X-12-ARIMA procedure to the four series. The two economic series have figures of merit showing that the fit is quite good. The F statistics for stable seasonality (month effect in a one-way ANOVA) are large, and the F statistics for moving seasonality (year effect

in a two-way ANOVA) are small, so that the seasonal components are judged present at 0.1%.

An important measure is the quarter or month of cyclical dominance, which is defined as the number of periods needed to make equal to 1, the ratio of the estimated irregular component to the estimated trend-cycle component. One quarter for GNP and 3 months for IIP are indication of good fits.

Under M statistics analysis the table reports the results of computing 11 sample quantities computed by the X-12-ARIMA program, meaning different characteristics of the seasonal fit. Each M statistic ranges from 0 to 3, and has acceptance values from 0 to 1. The Q statistic is an overall measure defined as a weighted average of the 11 M statistics. For the two economic series we find that no M statistics is larger than 1, and the Q statistics indicated that the seasonal fit is considered as accepted. A reference for these ideas is Lothian et al (1978).

Another way to study the quality of the seasonal fit is by means of the sliding spans analysis (Findley et al, 1990b). In the table we find that for the two economic series, the empirical values are either 0 or very small, and hence that the seasonal fit is deemed likely which is the most satisfactory category.

These analyses are also performed for the two rainfall series. The results for the F statistics, even when they differ in magnitude from those of the economic series, still lead to the qualification of seasonal present at 0.1%. The interpretation of the M and Q statistics are also not too negative: the series for Las Cejas is conditionally accepted, and that for La Cocha is rejected, in agreement with what we saw in other parts of this study. The month of cyclical dominance is 12 for both series, which is the worse possible value. Finally, the sliding spans analysis leads to the qualification of unlikely seasonal fits; we should point out that there are some reservations about the use of these methods in the case of an additive model (Findley et al, 1990).

Table 1 - Argentine's quarterly GNP, 1980-1996, one year ahead seasonal factors from X-12-ARIMA: bootstrap means and standard error estimates

a) Direct method

	Quarters				Medians
	I	II	III	IV	
1. One year ahead seasonal from X-12-ARIMA	94.746	103.512	101.232	100.578	100.905
2. Means over bootstrap replicates	94.748	103.430	101.160	100.650	100.905
3. Standard errors SE_{200}	0.390	0.476	0.474	0.442	0.458
4. Standard errors $SE_{200, 0.84}$	0.360	0.455	0.400	0.385	0.392
5. Standard errors $SE_{200, 0.90}$	0.371	0.488	0.421	0.406	0.413
6. Standard errors $SE_{200, 0.95}$	0.386	0.474	0.498	0.426	0.450

Medians of the estimates of SE	0.378	0.475	0.448	0.416	0.432
Standard deviations of the estimates of SE	0.012	0.012	0.039	0.021	0.027

b) AR(6) procedure

	Quarters				Medians
	I	II	III	IV	
1. One year ahead seasonal from X-12-ARIMA	94.746	103.512	101.232	100.578	100.905
2. Means over bootstrap replicates	94.672	103.489	101.201	100.627	100.914
3. Standard errors SE_{200}	0.490	0.588	0.525	0.510	0.518
4. Standard errors $SE_{200, 0.84}$	0.440	0.570	0.550	0.475	0.512
5. Standard errors $SE_{200, 0.90}$	0.480	0.554	0.527	0.480	0.503
6. Standard errors $SE_{200, 0.95}$	0.517	0.568	0.489	0.480	0.503

Medians of the estimates of SE	0.485	0.569	0.526	0.480	0.508
Standard deviations of the estimates of SE	0.028	0.012	0.022	0.014	0.006

Table 2 - Argentine's monthly IIP, 1980-1996, one year ahead seasonal factors from X-12-ARIMA: bootstrap means and standard error estimates

a) Direct method

	Months												Medians
	1	2	3	4	5	6	7	8	9	10	11	12	
1. One year ahead seasonal from X-12-ARIMA	92.320	80.520	102.930	99.260	104.320	100.770	102.960	105.570	102.710	106.000	104.380	98.270	103.545
2. Means over bootstrap replicates	92.279	80.740	102.119	99.583	103.633	101.285	103.724	105.047	103.314	105.627	103.782	98.783	103.548
3. Standard errors SE_{200}	1.188	1.003	1.246	1.214	1.227	1.208	1.132	1.143	1.111	1.172	1.204	1.112	1.142
4. Standard errors $SE_{200, 0.84}$	1.110	0.820	1.275	1.120	1.245	1.035	1.130	1.075	1.005	1.025	1.115	1.090	1.058
5. Standard errors $SE_{200, 0.90}$	1.076	0.870	1.283	1.100	1.174	1.069	1.115	1.057	1.065	1.135	1.174	1.108	1.121
6. Standard errors $SE_{200, 0.95}$	1.137	0.891	1.240	1.225	1.188	1.380	1.030	1.079	1.185	1.201	1.240	1.094	1.193
Medians of the estimates of SE	1.123	0.880	1.260	1.167	1.208	1.138	1.123	1.077	1.088	1.153	1.189	1.101	1.132
Standard deviations of the estimates of SE	0.041	0.067	0.018	0.055	0.029	0.136	0.042	0.033	0.066	0.067	0.046	0.009	0.049

b) AR(7) procedure

	Months												Medians
	1	2	3	4	5	6	7	8	9	10	11	12	
1. One year ahead seasonal from X-12-ARIMA	92.320	80.520	102.930	99.260	104.320	100.770	102.960	105.570	102.710	106.000	104.380	98.270	103.545
2. Means over bootstrap replicates	92.397	80.823	102.036	99.531	103.555	101.356	103.672	104.907	103.546	105.473	103.743	98.865	103.644
3. Standard errors SE_{200}	1.094	1.043	1.373	1.332	1.405	1.261	1.193	1.265	1.379	1.235	1.284	1.183	1.260
4. Standard errors $SE_{200, 0.84}$	1.075	0.975	1.285	1.280	1.285	1.080	1.150	1.235	1.205	1.175	1.160	1.055	1.168
5. Standard errors $SE_{200, 0.90}$	1.174	0.991	1.377	1.279	1.307	1.045	1.119	1.287	1.154	1.186	1.229	1.115	1.170
6. Standard errors $SE_{200, 0.95}$	1.201	0.979	1.459	1.356	1.398	1.140	1.195	1.204	1.307	1.237	1.207	1.191	1.222
Medians of the estimates of SE	1.134	0.985	1.375	1.306	1.352	1.110	1.172	1.250	1.256	1.210	1.218	1.149	1.196
Standard deviations of the estimates of SE	0.053	0.027	0.062	0.039	0.054	0.082	0.032	0.031	0.087	0.028	0.045	0.055	0.038

Table 3 - Argentine's economic series, means $S_{aj}^*(.)$ and standard errors $se_{200}(S_{aj})$, X-12-ARIMA, direct method, compared with constant regression seasonals and their standard errors

a) Quarterly GNP, 1980-1996

Quarter	Seasonal factors		Standard errors	
	Means over bootstrap replicates	Regression	Bootstrap se_{200} in (5)	Regression
I	94.748	94.008	0.390	0.477
II	103.430	101.969	0.476	0.483
III	101.160	102.081	0.474	0.489
IV	100.650	102.030	0.442	0.495

b) Monthly IIP, 1980-1996

Month	Seasonal factors		Standard errors	
	Means over bootstrap replicates	Regression	Bootstrap se_{200} in (5)	Regression
1	92.279	93.514	1.188	0.956
2	80.740	80.161	1.003	0.958
3	102.119	100.007	1.246	0.960
4	99.583	99.131	1.214	0.961
5	103.633	101.980	1.227	0.963
6	101.285	99.199	1.208	0.965
7	103.724	102.873	1.132	0.967
8	105.047	105.108	1.143	0.969
9	103.314	105.013	1.111	0.971
10	105.627	107.171	1.172	0.973
11	103.782	105.010	1.204	0.975
12	98.783	101.059	1.112	0.977

Table 4 - Las Cejas monthly rainfall, 1975-1989, one year ahead seasonal factors from X-12-ARIMA: bootstrap means and standard error estimates

a) Direct method

	Months												Medians
	1	2	3	4	5	6	7	8	9	10	11	12	
1. One year ahead seasonal from X-12-ARIMA	57.990	72.990	46.790	-10.440	-52.720	-43.900	-53.680	-51.490	-36.230	-26.400	16.810	78.960	-4.795
2. Means over bootstrap replicates	67.893	82.036	40.589	-9.432	-54.120	-46.056	-55.534	-53.787	-40.894	-25.933	22.173	71.632	-1.880
3. Standard errors SE_{200}	16.134	14.349	14.845	11.560	11.775	12.787	12.618	13.370	12.047	12.696	14.387	11.991	12.371
4. Standard errors $SE_{200, 0.84}$	13.435	13.510	12.015	9.830	9.895	11.270	10.625	11.620	10.890	11.245	11.295	10.170	11.068
5. Standard errors $SE_{200, 0.90}$	14.150	13.920	11.732	10.445	10.230	11.849	11.927	11.927	11.693	10.917	11.268	11.899	11.480
6. Standard errors $SE_{200, 0.95}$	14.854	14.550	12.599	10.970	10.286	12.574	12.845	12.216	11.991	11.094	13.44	11.739	11.865
Medians of the estimates of SE	14.502	14.134	12.307	10.707	10.258	12.212	12.272	12.071	11.842	11.170	12.369	11.819	11.672
Standard deviations of the estimates of SE	0.996	0.401	1.223	0.639	0.725	0.601	0.865	0.662	0.462	0.707	1.359	0.744	0.481

b) AR(12) procedure

	Months												Medians
	1	2	3	4	5	6	7	8	9	10	11	12	
1. One year ahead seasonal from X-12-ARIMA	57.990	72.990	46.790	-10.440	-52.720	-43.900	-53.680	-51.490	-36.230	-26.400	16.810	78.960	-4.795
2. Means over bootstrap replicates	69.762	82.372	41.718	-10.110	-54.010	-47.328	-53.208	-52.782	-40.749	-30.408	22.742	70.138	-3.833
3. Standard errors SE_{200}	14.600	14.395	14.342	13.499	13.582	14.057	12.342	13.140	13.715	12.700	16.276	13.857	13.786
4. Standard errors $SE_{200, 0.84}$	13.920	12.670	12.400	13.305	11.880	11.965	12.115	11.260	11.600	12.20	13.910	12.855	12.638
5. Standard errors $SE_{200, 0.90}$	13.935	12.574	12.664	13.522	13.112	12.820	11.689	12.367	11.603	12.356	14.423	13.986	13.171
6. Standard errors $SE_{200, 0.95}$	13.872	12.769	14.213	13.550	13.438	14.307	11.246	12.207	12.653	12.334	15.699	13.298	12.976
Medians of the estimates of SE	13.928	12.719	13.438	13.510	13.275	13.438	11.902	12.287	12.128	12.388	15.061	13.578	13.073
Standard deviations of the estimates of SE	0.300	0.750	0.879	0.096	0.670	0.948	0.419	0.669	0.876	0.146	0.950	0.453	0.418

Table 5 – La Cocha monthly rainfall, 1982-1997, one year ahead seasonal factors from X-12-ARIMA: bootstrap means and standard error estimates

a) Direct method

	Months												Medians
	1	2	3	4	5	6	7	8	9	10	11	12	
1. One year ahead seasonal from X-12-ARIMA	151.597	60.295	50.552	-42.848	-9.776	-55.899	-71.133	-71.530	-55.469	-25.306	-17.846	91.006	-21.576
2. Means over bootstrap replicates	134.472	67.066	31.899	-32.199	-11.240	-52.615	-67.510	-68.620	-53.334	-12.433	-9.510	76.116	-10.972
3. Standard errors SE_{200}	32.104	27.556	24.408	24.069	20.543	25.546	30.646	28.751	23.212	19.994	26.275	25.152	24.182
4. Standard errors $SE_{200, 0.84}$	19.855	21.260	18.760	21.785	16.330	17.750	18.970	20.515	20.145	18.245	20.690	19.550	19.848
5. Standard errors $SE_{200, 0.90}$	20.265	22.695	19.626	22.430	18.105	18.175	19.446	20.913	19.368	18.927	21.537	19.083	19.226
6. Standard errors $SE_{200, 0.95}$	21.277	22.787	21.036	22.781	18.997	20.976	19.921	21.921	20.711	18.541	24.845	20.043	20.377
Medians of the estimates of SE	20.771	22.741	20.331	22.605	18.551	19.575	19.684	21.417	20.428	18.734	23.191	19.796	20.112
Standard deviations of the estimates of SE	5.101	2.377	2.151	0.833	1.524	3.107	4.862	3.346	1.440	0.662	2.300	2.445	1.934

b) AR(12) procedure

	Months												Medians
	1	2	3	4	5	6	7	8	9	10	11	12	
1. One year ahead seasonal from X-12-ARIMA	151.597	60.295	50.552	-42.848	-9.776	-55.899	-71.133	-71.530	-55.469	-25.306	-17.846	91.006	-21.576
2. Means over bootstrap replicates	129.928	70.261	34.261	-29.135	-13.013	-54.149	-70.308	-71.341	-54.833	-8.816	-7.696	76.239	-8.256
3. Standard errors SE_{200}	25.561	24.883	24.198	22.013	18.947	20.610	19.637	20.973	19.504	20.683	21.663	21.845	21.173
4. Standard errors $SE_{200, 0.84}$	19.005	19.150	19.240	19.225	19.030	16.670	16.980	18.425	17.565	18.865	21.065	20.355	19.610
5. Standard errors $SE_{200, 0.90}$	19.532	20.320	20.476	19.282	17.964	18.986	17.980	20.289	19.115	18.397	20.242	20.480	19.678
6. Standard errors $SE_{200, 0.95}$	20.562	20.258	25.322	21.395	18.350	18.559	18.857	20.894	18.915	18.866	21.100	19.891	19.403
Medians of the estimates of SE	20.047	20.289	22.337	20.339	18.648	18.773	18.418	20.591	19.015	18.866	21.083	20.417	19.644
Standard deviations of the estimates of SE	2.599	2.203	2.521	1.245	0.439	1.403	0.991	1.028	0.730	0.876	0.507	0.728	0.704

Table 6 – Rainfall series, means S_{aj}^* (.) and standard errors $se_{200}(S_{aj}^*)$, X-12-ARIMA, direct method, compared with constant regression seasonals and their standard errors

a) Las Cejas, monthly, 1975-1989

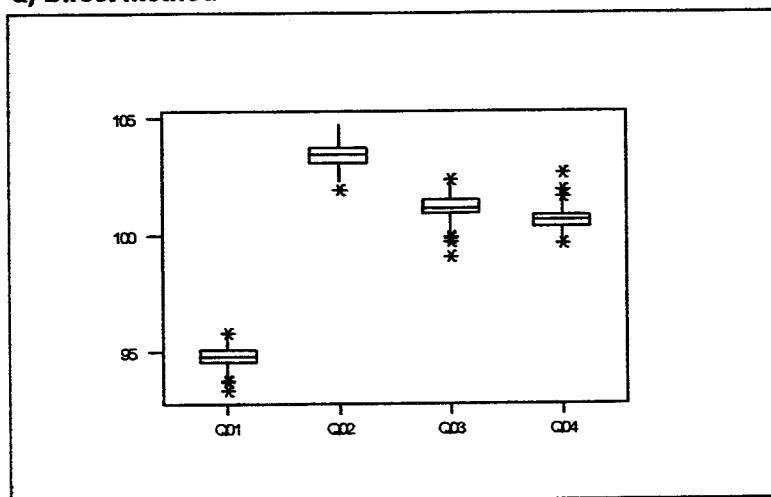
Month	Seasonal factors		Standard errors	
	Means over bootstrap replicates	Regression	Bootstrap se_{200} in (5)	Regression
1	67.893	102.503	13.435	12.206
2	82.036	89.794	13.510	12.233
3	40.589	84.610	12.015	12.260
4	-9.432	-6.486	9.830	12.286
5	-54.120	-55.483	9.895	12.314
6	-46.056	-57.368	11.270	12.341
7	-55.534	-59.863	10.625	12.369
8	-53.787	-58.907	11.620	12.397
9	-40.894	-40.234	10.890	12.425
10	-25.933	-20.502	11.245	12.453
11	22.173	20.104	11.295	12.482
12	71.632	96.549	10.170	12.510

b) Las Cocha, monthly, 1982-1997

Month	Seasonal factors		Standard errors	
	Means over bootstrap replicates	Regression	Bootstrap se_{200} in (5)	Regression
1	134.472	123.468	19.655	17.451
2	67.066	83.284	21.260	17.487
3	31.899	81.662	18.760	17.522
4	-32.199	8.768	21.785	17.559
5	-11.240	-15.371	16.330	17.595
6	-52.615	-42.735	17.750	17.632
7	-67.510	-58.069	18.970	17.669
8	-68.620	-63.286	20.515	17.706
9	-53.334	-45.931	20.145	17.744
10	-12.433	-20.593	18.245	17.782
11	-9.510	29.982	20.690	17.820
12	76.116	80.064	19.550	17.858

Figure 1 – Argentine's quarterly GNP, 1980-1996, one year ahead seasonal factors from X-12-ARIMA: box plots of bootstrap replicates

a) Direct method



b) AR(6) procedure

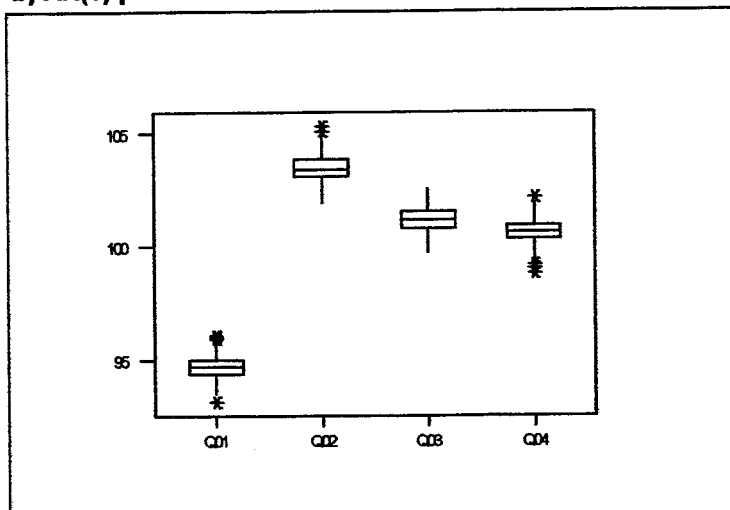
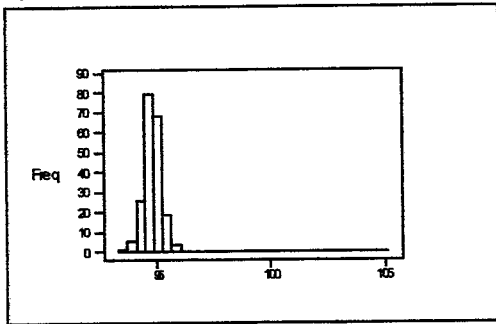
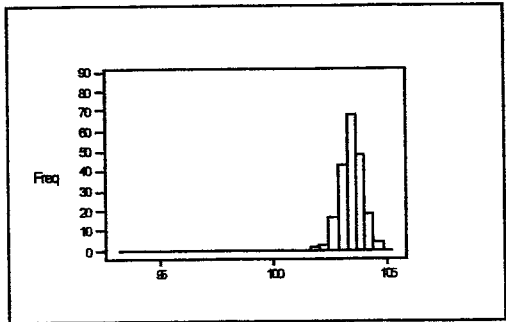


Figure 2 – Argentine's quarterly GNP, 1980-1996, one year ahead seasonal factors from X-12-ARIMA: histograms of bootstrap replicates by direct method, common 32 class intervals

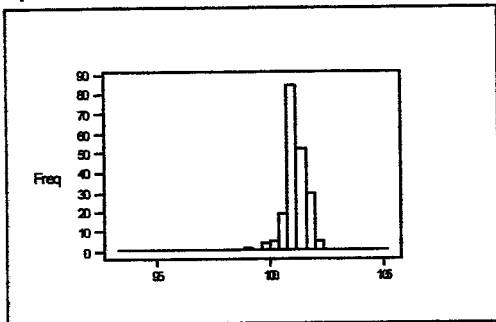
a) Quarter I



b) Quarter II



c) Quarter III



d) Quarter IV

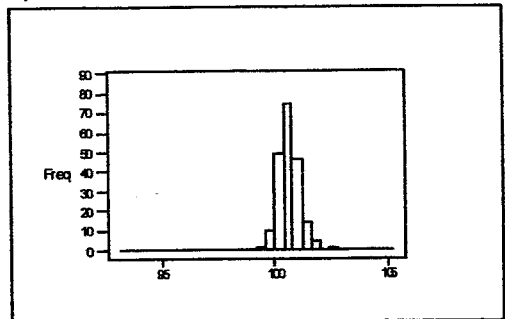
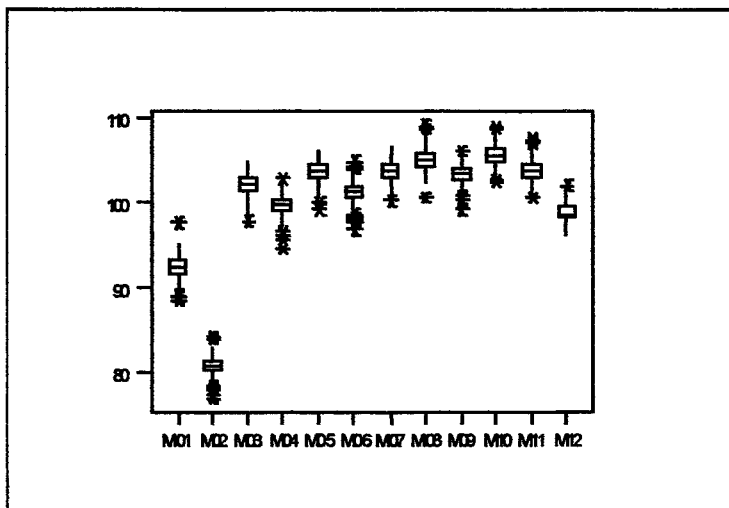


Figure 3 – Argentine's monthly IIP, 1980-1996, one year ahead seasonal factors from X-12-ARIMA: boxplots of bootstrap replicates

a) Direct method



b) AR(7) procedure

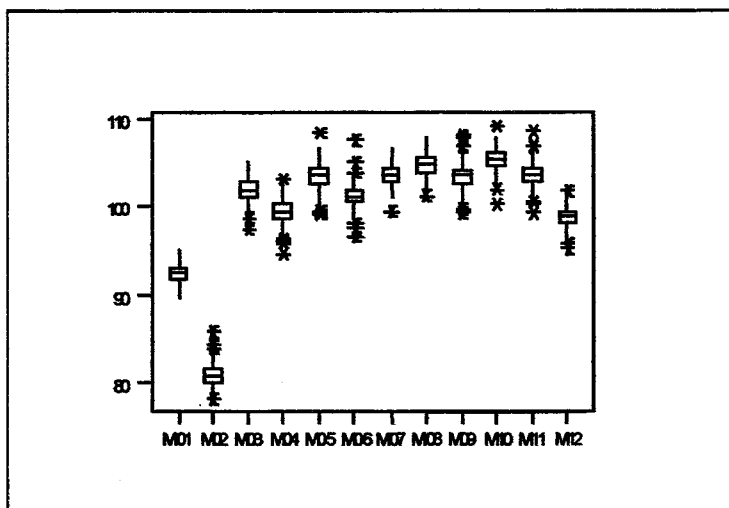
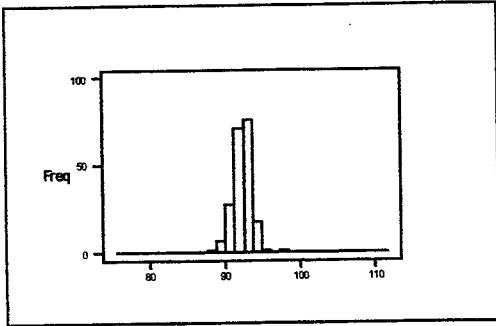
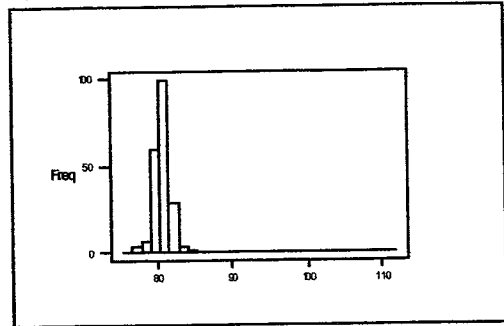


Figure 4 – Argentine's monthly IIP, 1980-1996, one year ahead seasonal factors from X-12-ARIMA: histograms of bootstrap replicates by direct method, common 32 class intervals (continued)

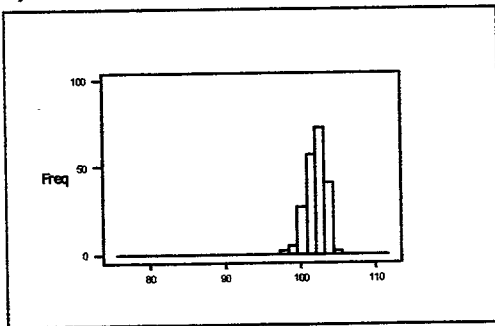
a) January



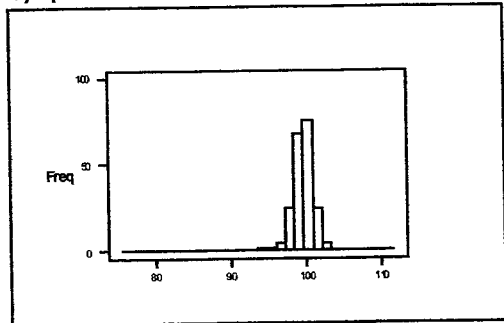
b) February



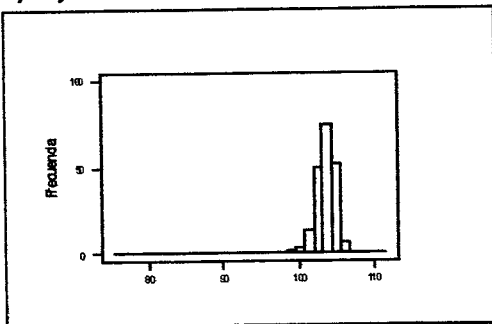
c) March



d) April



e) May



f) June

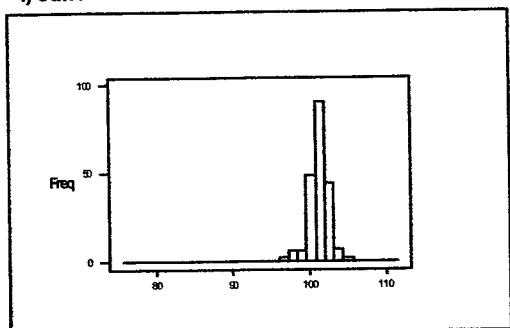
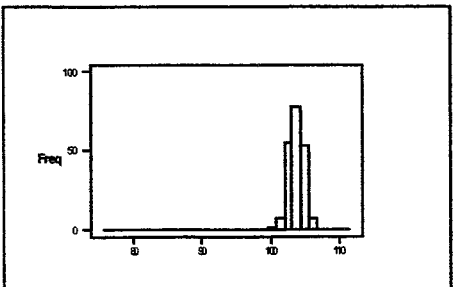
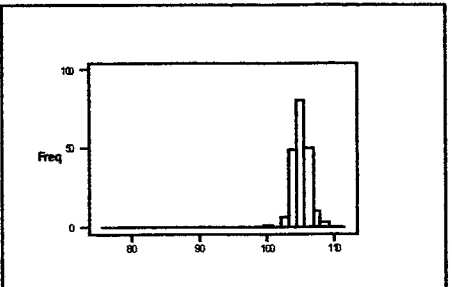


Figure 4 – Argentine's monthly IIP, 1980-1996, one year ahead seasonal factors from X-12-ARIMA: histograms of bootstrap replicates by direct method, common 32 class intervals (end)

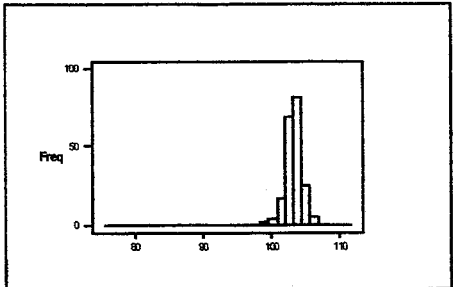
g) July



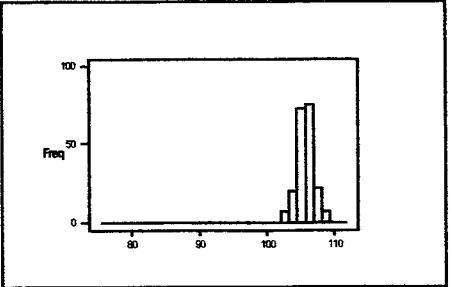
h) August



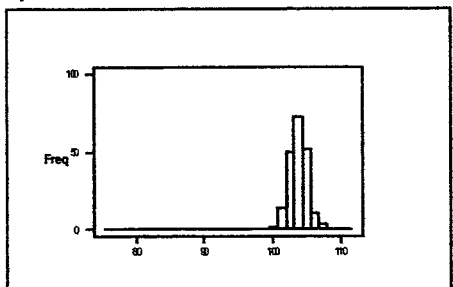
i) September



j) October



k) November



l) December

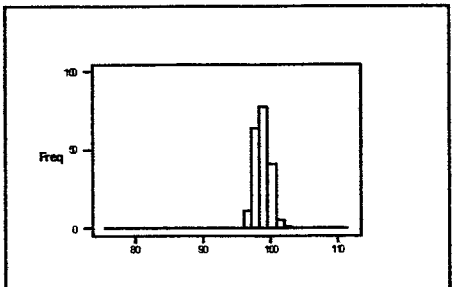
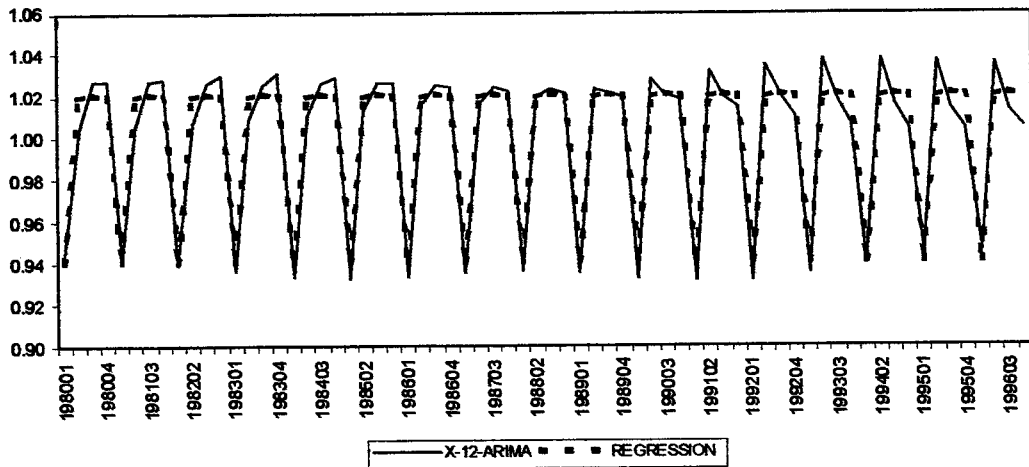


Figure 5 – Argentine's economic series, seasonal factors from X-12-ARIMA compared with constant regression seasonal

a) Quarterly GNP, 1980-1996



b) Monthly IIP, 1980-1996

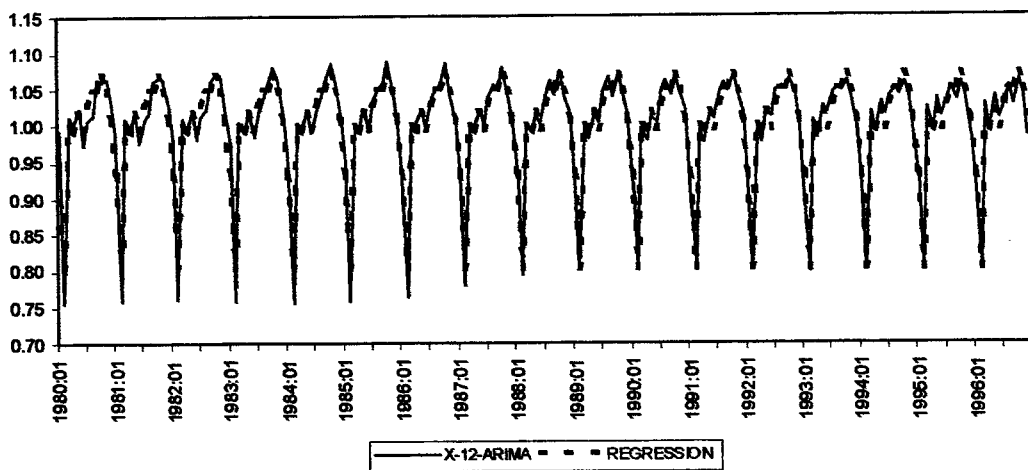
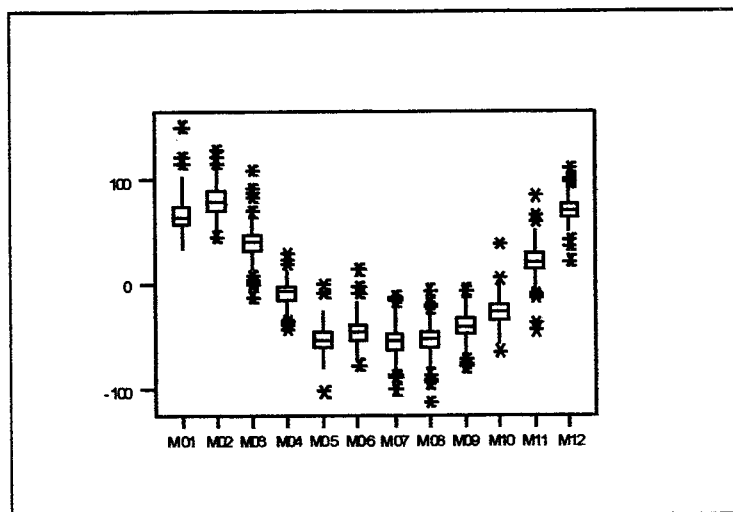


Figure 6 – Las Cejas monthly rainfall, 1975-1989, one year ahead seasonal factors from X-12-ARIMA: boxplots of bootstrap replicates

a) Direct method



b) AR(12) procedure

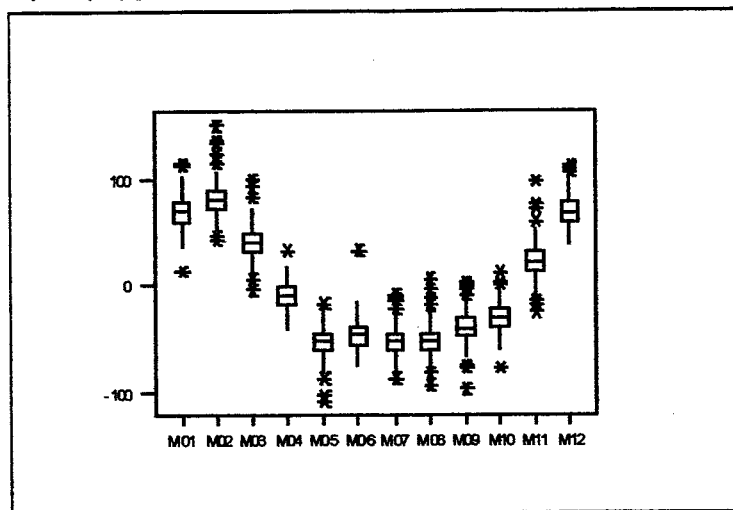
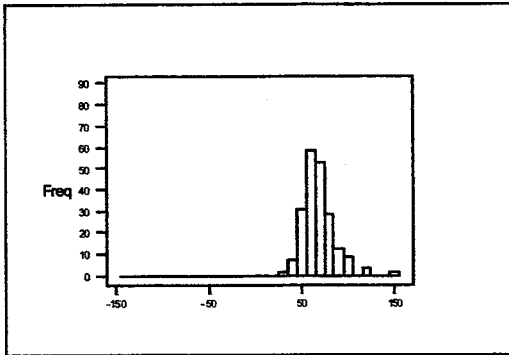
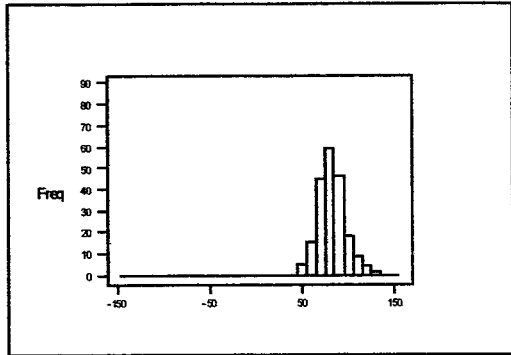


Figure 7 – Las Cejas monthly rainfall, 1975-1989, one year ahead seasonal factors from X-12-ARIMA: histograms of bootstrap replicates by direct method, common 32 class intervals (continued)

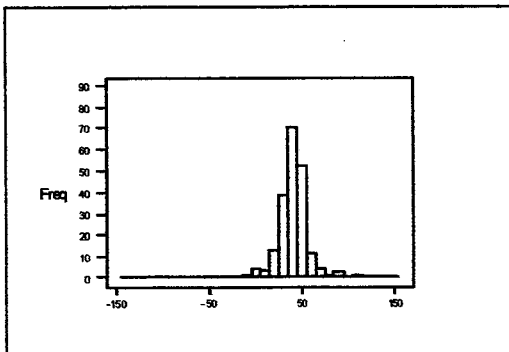
a) January



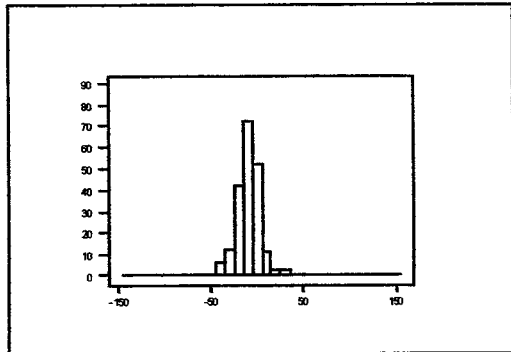
b) February



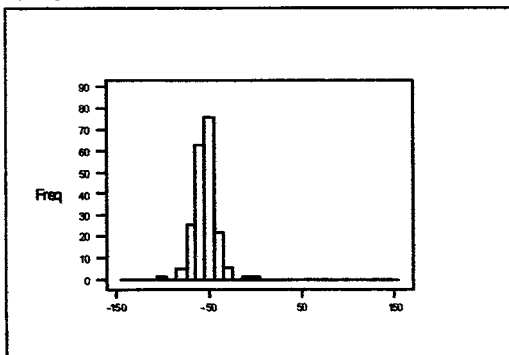
c) March



d) April



e) May



f) June

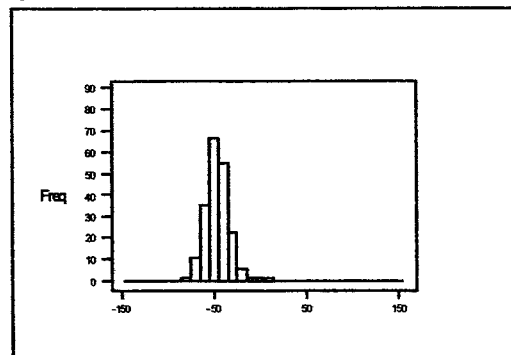
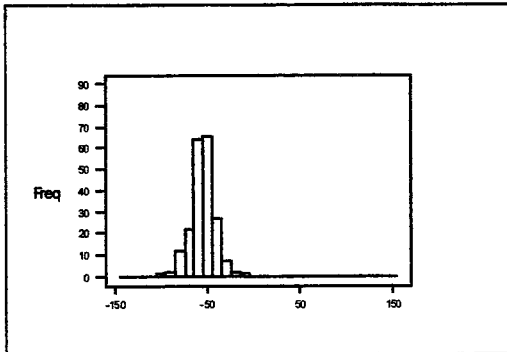
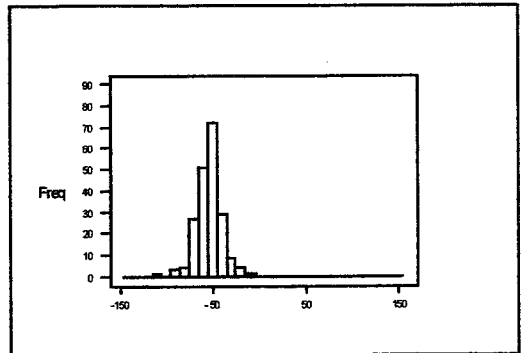


Figure 7 – Las Cejas monthly rainfall, 1975-1989, one year ahead seasonal factors from X-12-ARIMA: histograms of bootstrap replicates by direct method, common 32 class intervals (end)

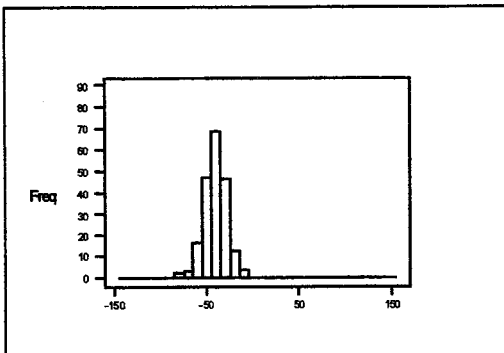
g) July



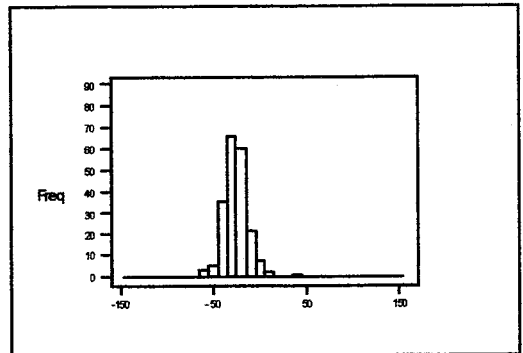
h) August



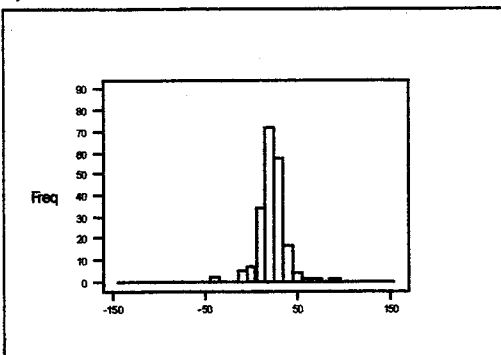
i) September



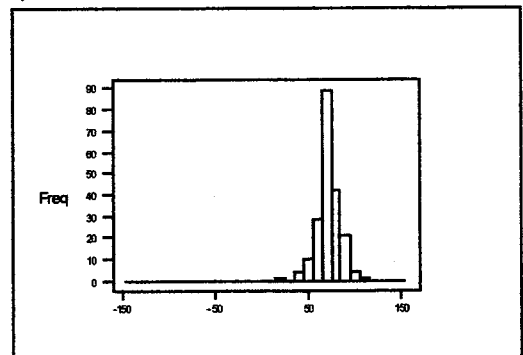
j) October



k) November

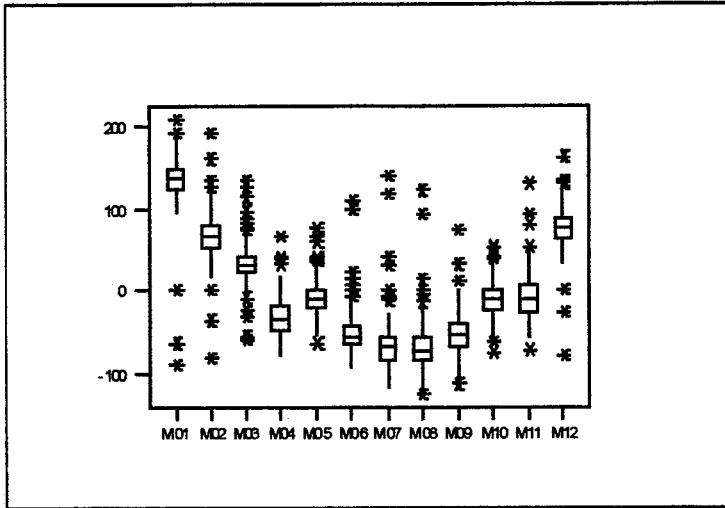


l) December



**Figure 8 – La Cocha rainfall, 1982-1997, one year ahead seasonal factors from X-12-ARIMA:
boxplots of bootstrap replicates**

a) Direct method



b) AR(12) procedure

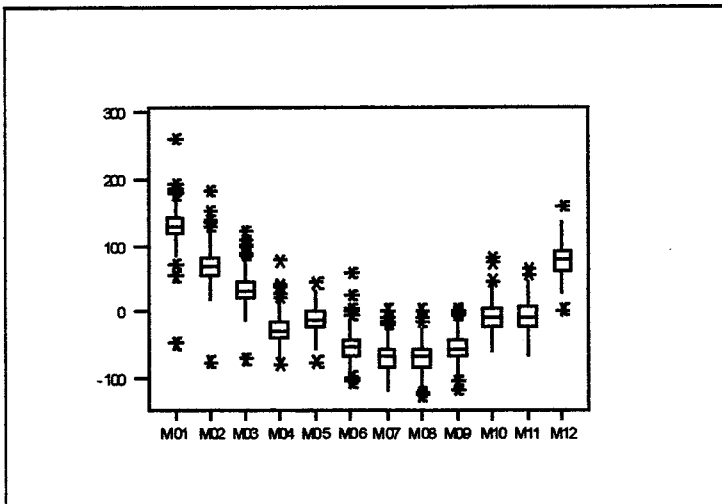
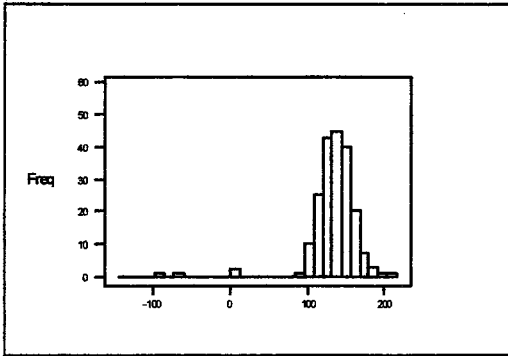
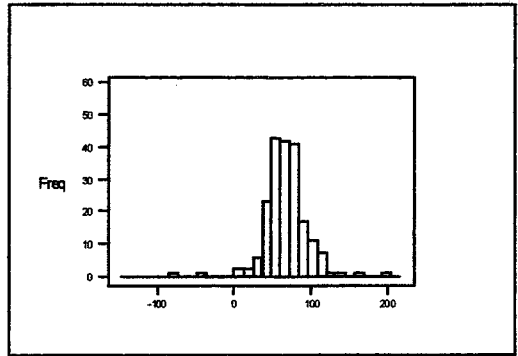


Figure 9 – La Cocha rainfall, 1982-1997, one year ahead seasonal factors from X-12-ARIMA: histograms of bootstrap replicates by direct method, common 32 class intervals (continued)

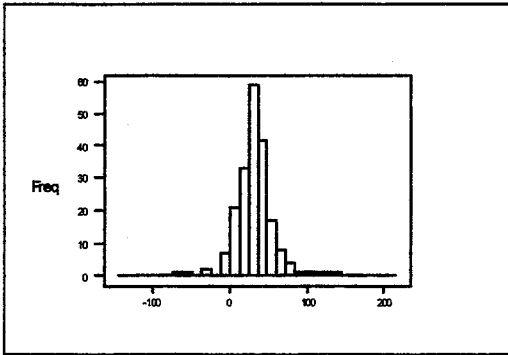
a) January



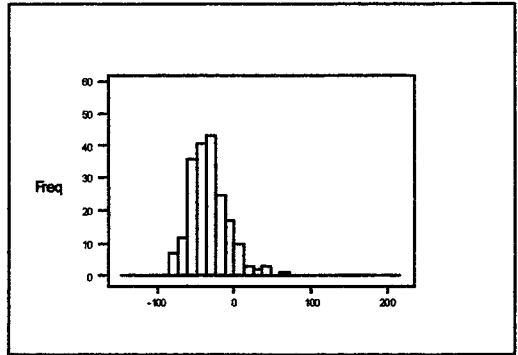
b) February



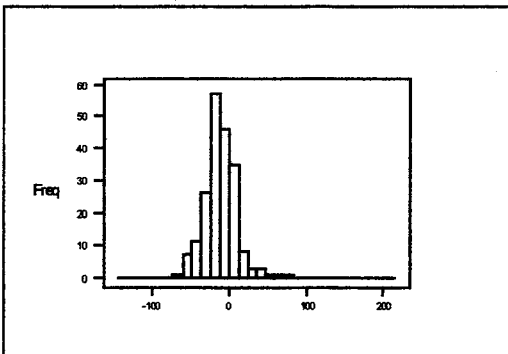
c) March



d) April



e) May



f) June

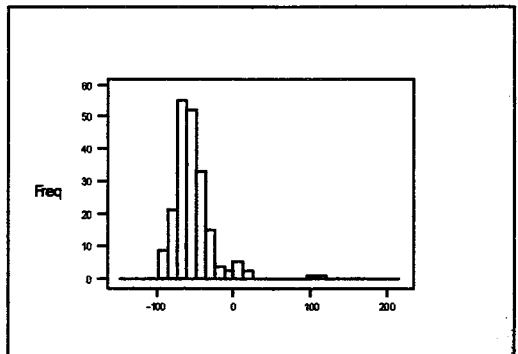
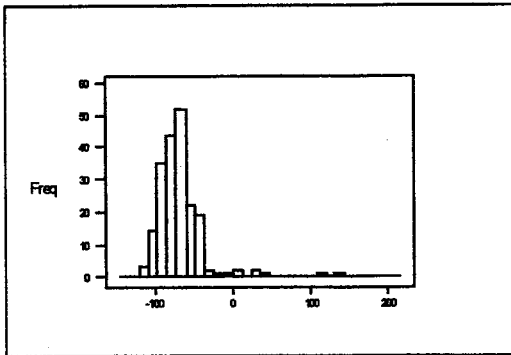
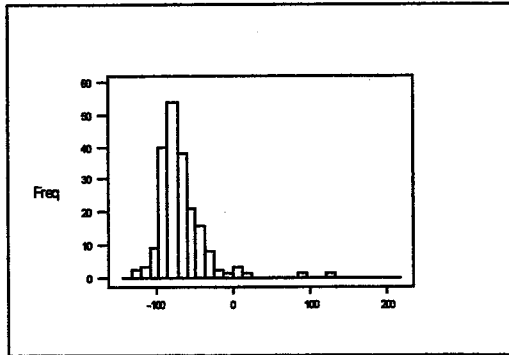


Figure 9 – La Cocha rainfall, 1982-1997, one year ahead seasonal factors from X-12-ARIMA: histograms of bootstrap replicates by direct method, common 32 class intervals (end)

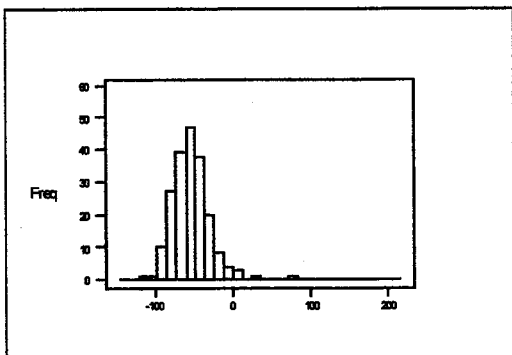
g) July



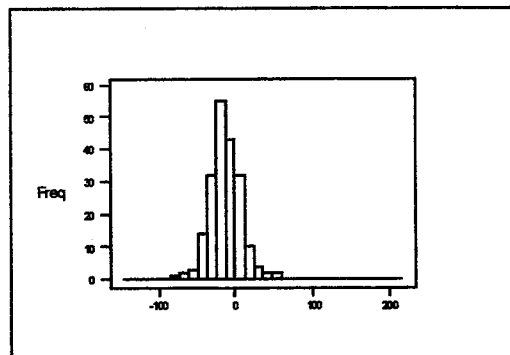
h) August



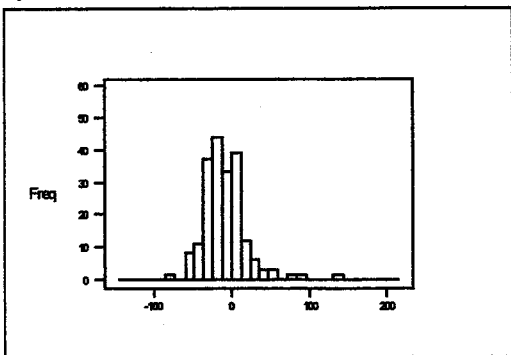
i) September



j) October



k) November



l) December

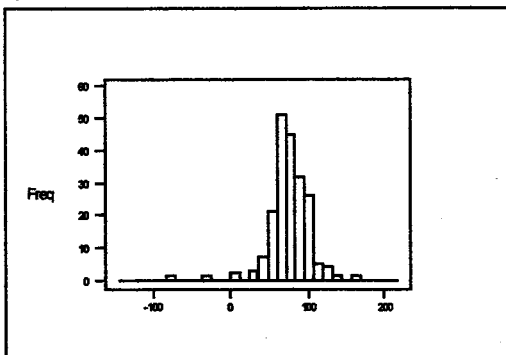
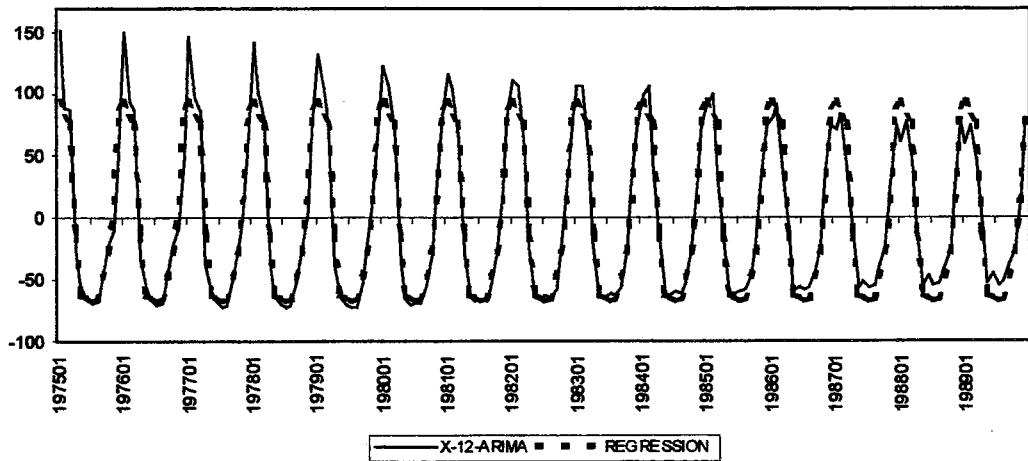
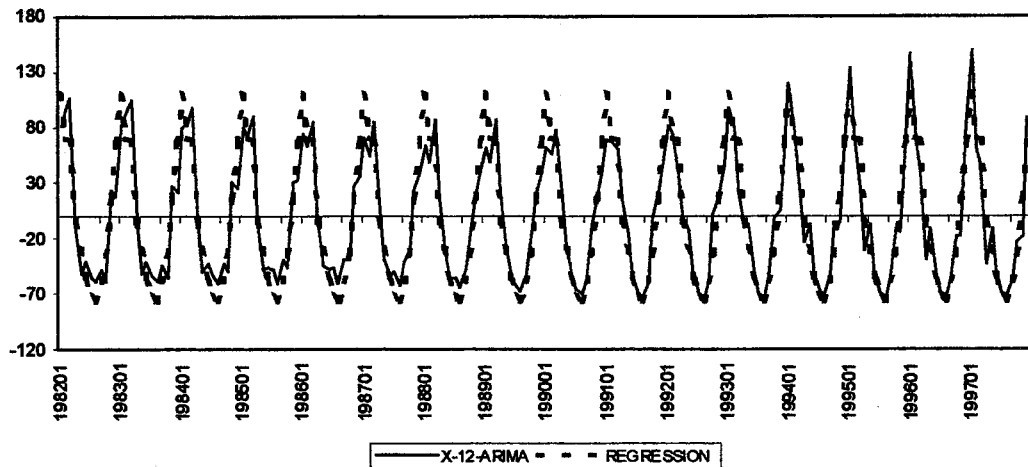


Figure 10 – Rainfall series, seasonal factors from X-12-ARIMA compared with constant regression seasonal

a) Las Cejas, monthly, 1975-1989

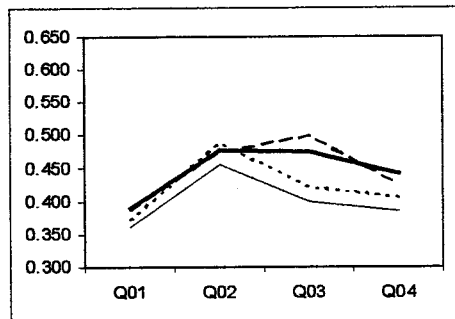


b) La Cocha, monthly, 1981-1997

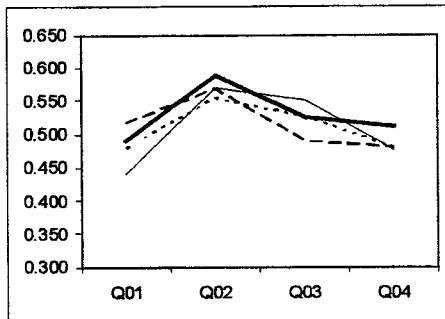


**Figure 11 – Argentine's economic series, one year ahead seasonal factors from X-12-ARIMA:
graphical comparison of four standard error estimates**

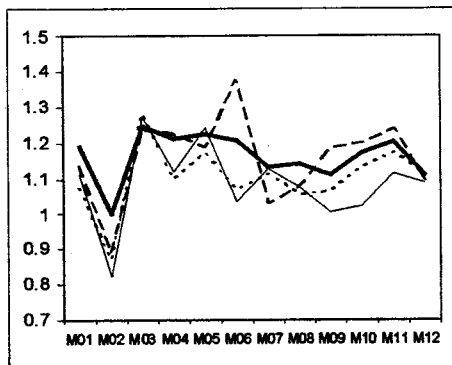
a) Quarterly GNP, 1980-1996, direct method



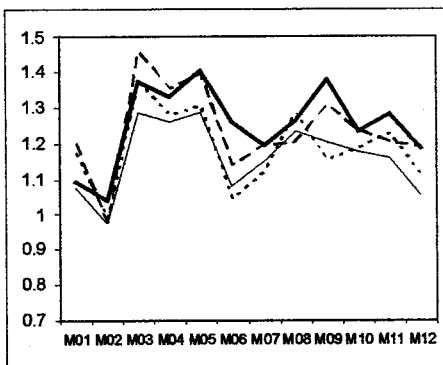
b) Quarterly GNP, 1980-1996, AR(6) procedure



a) Monthly IIP, 1980-1996, direct method



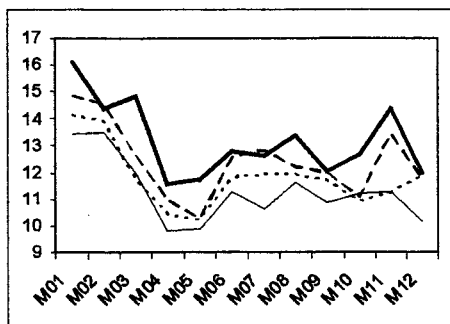
b) Monthly IIP, 1980-1996, AR(7) procedure



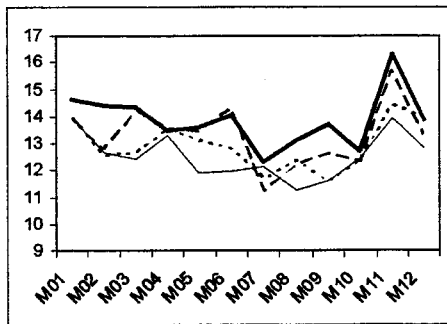
——— 3.Standard errors se_{200}
 - - - - 4.Standard errors $se_{200,0.94}$
 - - - - 5.Standard errors $se_{200,0.90}$
 - - - - 6.Standard errors $se_{200,0.85}$

**Figure 12 – Argentine’s rainfall series, one year ahead seasonal factors from X-12-ARIMA:
graphical comparison of four standard error estimates**

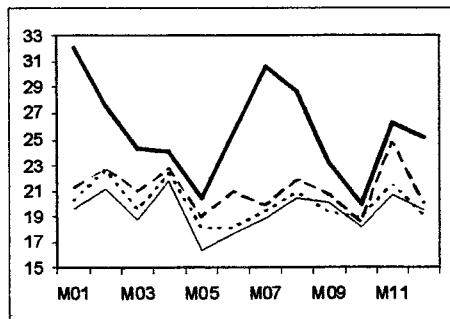
a) Las Cejas, monthly 1975-1989, direct method



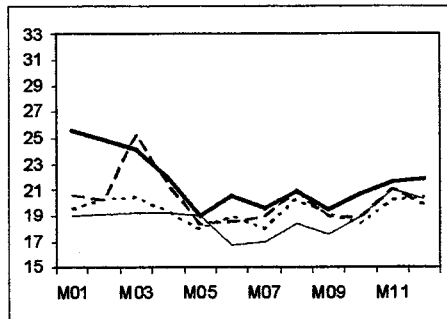
b) Las Cejas, monthly 1975-1989, AR(12) procedure



c) La Cocha, monthly 1982-1997, direct method



d) La Cocha, monthly 1982-1997, AR(12) procedure



——— 3. Standard errors se_{200}
 - - - - 4. Standard errors $se_{200,0.24}$
 - · - · 5. Standard errors $se_{200,0.50}$
 - - - - 6. Standard errors $se_{200,0.85}$

Bibliography

- BROCKWELL, P.J. and R.A. DAVIS, (1991a). *Time Series: Theory and Methods, 2nd Edition*, New York: Springer Verlag.
- BROCKWELL, P.J. and R. A. DAVIS, (1991b). *ITSM: An Interactive Time Series Modelling Package for the PC*, New York: Springer Verlag.
- CLEVELAND, W. P. and G. C. TIAO, (1976). Decomposition of seasonal time series: a model for the Census X-11 program, *J. of the American Statistical Association*, Vol. 71, No 355, 581-587.
- EFRON, B. and F. TIBSHIRANI, (1993). *An Introduction to the Bootstrap*, New York: Chapman and Hall.
- FINDLEY, D. F. and B. C. MONSELL, (1989). Reg-ARIMA model preprocessing for seasonal adjustment. In A. C. Singh and P. Whitridge, editors, *Analysis of Data in Time Series*, pages 117-123, Ottawa: Statistics Canada.
- FINDLEY, D. F. and B. MONSELL, (1990a). Standard errors for seasonal adjustment: A resampling approach, *Proceedings of the Business and Economics Statistics Section*, pages 455-460, Alexandria, Virginia, American Statistical Association.
- FINDLEY D.F., B.C. MONSELL, H.B. SHULMAN, and M.G. PUGH, (1990b). Sliding spans diagnostics for seasonal and related adjustment, *J. of the American Statistical Association*, Vol. 85, 345-355
- FINDLEY, D. F., B., C. MONSELL, M. C. OTTO and B.CHEN, (1998). New capabilities and methods of the X-12-ARIMA seasonal adjustment program, *J. of Business and Economic Statistics*, Vol. 16, No. 2, 127-176.
- LI, H. and G. S. MADDALA, (1996). Bootstrapping time series models, *Econometric Reviews*, Vol. 15, No. 2, 115-158
- LOTHIAN, J. and M. MORRY, (1978). A set of quality control statistics for the X11-ARIMA seasonal adjustment method, *Working paper 78-10-005*, Methodology Branch, Seasonal Adjustment and Time Series Analysis Staff, Statistics Canada, Ottawa.
- MENTZ R, M. D'URSO , N. JARMA and G. MENTZ, (2000). Cyclical components of local rainfall data, *International Journal of Climatology*, Vol. 20, Number 2, pages 211-228.
- PFEFFERMANN, D., (1994). A general method for estimating the variances of X-11 seasonally adjusted estimators, *J. of Time Series Analysis*, Vol. 15, No. 1, 85-116.
- SHISKIN, J., A. H. YOUNG and J. C. MUSGRAVE, (1967). The X-11 variant of the Census Method II seasonal adjustment. Technical Paper No. 15, U. S. Bureau a of the Census, Washington.
- U.S. BUREAU OF THE CENSUS, (1999). *X-12-ARIMA Reference Manual*, Final Version 0.2.5 Washington D.C.: U.S. Bureau of the Census.
- Wallis, K. F., (1974). Seasonal adjustment and relations between variables, *J. of the American Statistical Association*, Vol. 69, No. 345, 18-31.

Acknowledgments

The authors thank Marcela D'Urso for help in the initial phase of this study and in the preparation of the Appendix. They thank the comments by Professor A.J. Viollaz, and a conversation held with Professor Lutz Kilian at the University of Michigan. Reports by two referees were very useful to improve the contents and presentation.

REVISTA BRASILEIRA DE ESTATÍSTICA - RBEs

POLÍTICA EDITORIAL

A Revista Brasileira de Estatística - RBEs publica trabalhos relevantes em Estatística Aplicada, não havendo limitação no assunto ou matéria em questão. Como exemplos de áreas de aplicação citamos as áreas de advocacia, ciências físicas e biomédicas, criminologia, demografia, economia, educação, estatísticas governamentais, finanças, indústria, medicina, meio ambiente, negócios, políticas públicas, psicologia e sociologia, entre outras. A RBEs publicará, também, artigos abordando os diversos aspectos de metodologias relevantes para usuários e produtores de estatísticas públicas, incluindo planejamento, avaliação e mensuração de erros em censos e pesquisas, novos desenvolvimentos em metodologia de pesquisa, amostragem e estimação, imputação de dados, disseminação e confiabilidade de dados, uso e combinação de fontes alternativas de informação e integração de dados, métodos e modelos demográficos e econométricos. Os artigos submetidos devem ser inéditos e não devem ter sido submetidos simultaneamente a qualquer outro periódico.

O periódico tem como objetivo a apresentação de artigos que permitam fácil assimilação por membros da comunidade em geral. Os artigos devem incluir aplicações práticas como assunto central, com análises estatísticas exaustivas e apresentadas de forma didática. Entretanto, o emprego de métodos inovadores, apesar de ser incentivado, não é essencial para a publicação.

Artigos contendo exposição metodológica são também incentivados, desde que sejam relevantes para a área de aplicação pela qual os mesmos foram motivados, auxiliem na compreensão do problema e contenham interpretação clara das expressões algébricas apresentadas.

A RBEs tem periodicidade semestral e também publica artigos convidados e resenhas de livros, bem como incentiva a submissão de artigos voltados para a educação estatística.

Artigos em espanhol ou inglês só serão publicados caso nenhum dos autores seja brasileiro e nem resida no País.

Todos os artigos submetidos são avaliados quanto à qualidade e à relevância por dois especialistas indicados pelo Comitê Editorial da RBEs.

O processo de avaliação dos artigos submetidos é do tipo 'duplo cego', isto é, os artigos são avaliados sem identificação de autoria e os comentários dos avaliadores também são repassados aos autores sem identificação.

INSTRUÇÃO PARA SUBMISSÃO DE ARTIGOS À RBEs

O processo editorial da RBEs é eletrônico. Os artigos devem ser submetidos via e-mail para: hortega@ibge.gov.br

Após a submissão, o autor correspondente receberá um código para acompanhar o processo de avaliação do artigo. Caso não receba um aviso com este número no prazo de uma semana, fazer contato com a secretaria da revista no endereço:

Revista Brasileira de Estatística

IBGE – Diretoria de Pesquisas - Coordenação de Métodos e Qualidade

Av. República do Chile, nº 500, 10º andar

Centro, Rio de Janeiro – RJ

CEP: 20031-170

Tels.: 55 21 2142-0472

55 21 2142-4549

Fax: 55 21 2142-4802

INSTRUÇÕES PARA PREPARO DOS ORIGINAIS

Os originais entregues para publicação devem obedecer às normas seguintes.

1. Originais processados pelo editor de textos Word for Windows são preferidos. Entretanto, serão aceitos, também, originais processados em LaTeX desde que sejam encaminhados acompanhados de versões em pdf, conforme descrito no item 3 a seguir;
2. A primeira página do original (folha de rosto) deve conter o título do artigo, seguido do(s) nome(s) completo(s) do(s) autor(es), indicando-se, para cada um, a afiliação e endereço para correspondência. Agradecimentos a colaboradores e instituições, e auxílios recebidos, também, devem figurar nesta página;
3. No caso de a submissão não ser em Word for Windows, três arquivos do original devem ser enviados. O primeiro deve conter os originais no processador de texto utilizado (por exemplo, Latex). O segundo e terceiro devem ser no formato pdf, sendo um com a primeira página, como descrito no item 2, e outro contendo apenas o título, sem identificação do(s) autor(es) ou outros elementos que possam permitir a identificação da autoria;
4. A segunda página do original deve conter resumos em português e inglês (*abstract*), destacando os pontos relevantes do artigo. Cada resumo deve ser digitado seguindo o mesmo padrão do restante do texto, em um único parágrafo, sem fórmulas, com, no máximo, 150 palavras;
5. O artigo deve ser dividido em seções, numeradas progressivamente, com títulos concisos e apropriados. Todas as seções e subseções devem ser numeradas e receber título apropriado;

6. **Tratamentos algébricos exaustivos devem ser evitados ou alocados em apêndices;**
7. **A citação de referências no texto e a listagem final de referências devem ser feitas de acordo com as normas da ABNT;**
8. **As tabelas e gráficos devem ser precedidos de títulos que permitam perfeita identificação do conteúdo. Devem ser numeradas seqüencialmente (Tabela 1, Figura 3, etc.) e referidas nos locais de inserção pelos respectivos números. Quando houver tabelas e demonstrações extensas ou outros elementos de suporte, podem ser empregados apêndices. Os apêndices devem ter título e numeração, tais como as demais seções de trabalho; e**
9. **Gráficos e diagramas para publicação devem ser incluídos nos arquivos com os originais do artigo. Caso tenham que ser enviados em separado, devem ter nomes que facilitem a sua identificação e posicionamento correto no artigo (ex. Gráfico 1; Figura 3; etc.). É fundamental que não existam erros, quer no desenho, quer nas legendas ou títulos.**