

Estudo de Caso final: Ancova

Equipe 04

03 de Julho de 2017

Coordenador: Alessandro Cardoso

Relator: Bernardo Marques

Verificador: Danny Tonidandel

Monitor: Gustavo Vieira

1. Descrição do Problema

O experimento envolve comparações entre métodos aplicados em diferentes instâncias de problemas de roteamento, configurando o clássico problema do “caixeiro viajante”. A variabilidade decorrente das características de cada problema de teste é uma possível fonte de variação espúria quando não considerada na análise dos resultados. Para avaliar existência de diferenças significativas entre os algoritmos, bem como a influência dos possíveis covariantes, tais como tamanho da instância e tempo de convergência, será utilizada a análise de covariância ANCOVA.

Os parâmetros experimentais desejados são:

- Nível de significância: $\alpha = 0.05$;
- Tamanho de efeito de interesse prático: $\delta^* = 0.25$;
- Potência desejada: $(1 - \beta) = \pi \geq 0.85$.

2. Planejamento Experimental

A primeira etapa do experimento consistiu na geração dos dados experimentais a partir dos dois algoritmos de otimização. Busca-se a verificação de diferenças estatísticas significativas entre as configurações testadas. Uma análise exploratória qualitativa dos dados foi feita para complementar os testes.

Como não se tinha ideia a respeito do tamanho amostral necessário, foi realizado o cálculo pensando-se em utilizar o teste da análise de variância para avaliar a existência de diferenças significativas entre as classes de problemas. Embora a técnica permita analisar as médias e variâncias de observações de diferentes grupos, generalizando o teste-t para mais de dois deles [Campelo, 2015], ela é restrita apenas à indicação de existência ou não de diferenças entre os níveis avaliados, sem indicar quais níveis seriam diferentes [Montgomery and Runger, 2011], tampouco a influência de variáveis concomitantes. Neste ponto, evidencia-se a necessidade de uma análise de covariância.

3. Análise de Co-Variância (ANCOVA)

A análise de covariância é uma técnica — assim como a chamada “blocagem” ou pareamento (em testes-t) — bastante útil para a melhora da precisão de um experimento [Montgomery, 1984]. Em diversos aspectos é similar à análise de variância (ANOVA), porém permite ter controle sobre a influência do covariante nas variáveis dependentes. A covariável complementa o controle local e, obviamente, necessita estar correlacionada com a variável de resposta para que se possa fazer uso de tal análise. E quando a análise de variância é realizada com uma ou mais covariáveis, é usual chamar a análise de ANCOVA.

Na ANCOVA a variável dependente é contínua (e.g. tempo, velocidade, etc.), enquanto que a variável independente é normalmente categórica (e.g. “masculino/feminino”, “fumante/não-fumante” etc.). A análise de variância, por sua vez, poderá converter-se na ANCOVA quando for adicionado um covariante, que consiste em outra variável, que pode ser tanto categórica quanto contínua.

Existem aliás, duas razões principais para sair-se da ANOVA e passar a ser considerada a análise de covariância, i.e., razões técnicas para adicionar um covariante, quais sejam:

- Reduzir fatores de variabilidade inter-grupos, ou alcançar um nível maior de entendimento a partir da variância desconhecida;
- Isolar o efeito que ocorre quando o controle experimental não permite que o experimentador elimine, de maneira razoável, explicações alternativas para uma relação observada entre variáveis independentes e dependentes, o que é chamado de “confusão” *confounding*. Em outras palavras, esta variável é algo que pode estar influenciando o experimento mas que não está, a princípio, no modelo original, e é uma potencial fonte de vieses no experimento.

A ANCOVA, permite, portanto, um controle do erro experimental, aumentando sua precisão. Vale ressaltar que a técnica não é restrita apenas à indicação de existência ou não de diferenças entre os níveis avaliados.

3.1 outra subseção

3.2 Cálculo do tamanho amostral para o ANOVA

O cálculo do tamanho amostral para a técnica ANOVA pode ser feito iterativamente até encontrar o número n tal que:

$$F_{(1-\alpha)} = F_{\beta;\phi}, \quad (1)$$

em que ambas distribuições F têm $(a - 1)$ graus de liberdade no numerador e $a(n - 1)$ no denominador. O parâmetro de não-centralidade ϕ é dado por:

3.3 Tratamento e Validação dos Dados

4. Análise Estatística

4.1 Análise de Covariância

4.2 Validação das Premissas

Normalidade

Homocedasticidade

Independência

5. Discussão e Conclusões

Referências

Felipe Campelo. Lecture notes on design and analysis of experiments. <https://github.com/fcampelo/Design-and-Analysis-of-Experiments>, 2015. Version 2.11, Chapter 7; Creative Commons BY-NC-SA 4.0.

D. C. Montgomery. *Design and analysis of experiments*, volume 7. Wiley New York, 1984.

D. C. Montgomery and G. C. Runger. *Applied Statistics and Probability for Engineers*, volume 5. John Wiley and Sons, 2011.