bons-exemplos-teste-hipotese

Amanda Souza

23/05/2018

## Perguntas

I - Procure e escolha um artigo científico que em sua avaliação faz bom uso testes de hipótese. Poste nesta discussão uma descrição de:

1. Qual a pergunta que é respondida utilizando testes de hipótese no artigo? (Se houver várias tudo bem escolher uma). Quem é população e quem é amostra? Qual a hipótese nula no teste? Como o p-valor é calculado?
2. Como os autores do artigo reportam textualmente o resultado envolvendo o teste, tanto os detalhes estatísticos quanto as conclusões tiradas a partir desses detalhes? (aqui tudo bem transcrever)

Por hora, eu sugiro que evitem-se artigos com regressões. Não esqueça de postar um link para o artigo também.

## Respostas

O artigo escolhido possui o tema [An Empirical Study for Evaluating the Performance of Multi-cloud APIs](https://www.sciencedirect.com/science/article/pii/S0167739X17301802) e aborda o estudo de APIs para orquestração de carga de trabalho em várias plataformas de nuvem.

A pergunta principal do artigo aborda como o avaliar o desempenho de APIs implantadas em diversos tipos de nuvem quando comparado a APIs em uma plataforma específica, como Amazon Web Services e Microsoft Azure, para impulsionar decisões tecnológicas em aplicativos em nuvem que exigem desempenho máximo ao usar várias nuvens diferentes. A população estudada no artigo foi os dados das APIs das linguagens Java e Python nas clouds: Amazon Web Services e Microsoft Azure. A amostra consiste em um dataset de 39,722 dados provenientes das observações coletadas nas duas clouds em cinco dias diferentes.

No artigo são realizados alguns testes de hipótes, mas neste tópico serão definidas as hipóteses principais:

H0: a mediana de desempenho é a mesmo, independentemente do tipo de API na nuvem. H1: a mediana de desempenho é diferente para qualquer API na nuvem.

O p-value é calculado usando o teste de hipótese do Bonferroni, e os teste de Wilcoxon para calcular as diferenças entre Java e Python. A comparação é realizada pelo o valor mediano dos dados com um valor hipotético específico, que resultou em α = 0,00125 para aceitar ou rejeitar a hipótese nula. Esses cálculos são baseados na latência, memória e tempo de resposta das APIs de nuvem.

Os resultados dos testes para latência apresentaram que todos os fatores são representativos (p-valor ≤ 0,001) em α = 0,05. Isso sugere que o tempo de CPU não varia significativamente em dias diferentes. O teste de Wilcoxon confirma que as diferenças de CPU entre APIs multi-nuvem e específicas de plataforma são estatisticamente significativas (p-valor < 0,00125). A hipótese nula de latência foi rejeitada e a hipótese alternativa foi aceita, afirmando que o tempo médio de CPU varia de acordo com o tipo de API de nuvem usada.

Os resultados para memória, todos os fatores são representativos (p-valor <0,003) em α = 0,05. Isso sugere que o consumo de memória não varia significativamente em dias diferentes. No teste de Wilcoxon, as diferenças são estatisticamente significativas (valor de p-value < 0.0001). A hipótese nula de memória foi rejeitada e a hipótese alternativa foi aceita, afirmando que o consumo médio de memória varia de acordo com o tipo de API da nuvem usada.

No tempo de resposta, todos os fatores são representativos (p-valor ≤ 0,0152) em α = 0,05. Ao contrário dos outros dois indicadores de desempenho, dois ensaios foram representativos (26/06/16 e 28/06/2016). Isso sugere que esses dias afetam o tempo de resposta. O tempo de resposta foi semelhante para APIs multi-nuvem e API específica, e a diferença entre elas não foram consideradas significativas (p-valor > 0,375). A hipótese nula para o tempo de resposta foi aceita e a hipótese alternativa foi rejeitada, afirmando que o tempo de resposta não varia com o tipo de API da nuvem usada.