My wonderful presentation

Alexey Gumirov

Contents

- Código genético - Degenerado
- Preferência por códons específicos
  - Viés de uso de codon (*codon usage bias*)
  - Comum no genoma mitocondrial
- Possíveis explicações
  - Maximizar a transcrição de seus genes (e.g. mais A-T)
  - Priorizar os códons do mitogenoma (22 tRNAs)
- Clado escolhido: Primates

Perguntas

Pergunta Biológica

Quais aminoácidos apresentam *codon usage bias* (se é que algum apresenta)?

Pergunta Quantitativa

Os valores observados na contagem de códons sinônimos mitocondriais para um dado aminoácido são diferentes dos valores esperados caso não haja preferência por nenhum códon?

Estudo observacional

- Amostragem: 199 espécies - Teorema Central do Limite
- Não usaremos todas as variáveis medidas
  - **Variável Dependente**:
    - Codon count
  - **Variável Independente**:
    - Codon
- Unidade experimental: Espécie
  - Mitogenoma completo da espécie
- Logo, minha pergunta e possíveis conclusões estão **restritos à mitocôndrias de primatas** (amostra é representativa dessa população).

Distribuição gaussiana é comum para variáveis que:

1. Sejam influenciadas por um grande numero de fatores, sem que nenhum deles seja dominante (parece ser o caso)
2. Não sejam constritas em sua variação (já que estamos trabalhando com valores absolutos, pode ser q seja)

- Teorema do limite central: Independentemente da distribuição de uma variável na população, as **médias de amostras aleatórias** retiradas desta população se distribuirão em uma curva que tenderá à normal conforme o tamando da amostra aumenta.

- Alpha (limiar significância): É a probabilidade de erro Tipo I (falso positivo). Quanto mais estrito, menor o poder.
- Beta: É a probabilidade de erro Tipo II.

– O poder é a **chance de um experimento detectar uma diferença quando ela existe (1 - beta).**

Teste estatístico:

- Teste para cada aminoácido
  - Uma variável qualitativa independente (codons) em cada teste
  - Essa variável qualitativa única pode ter duas ou mais categorias
  - A variável dependente é a contagem dos códons
  - Se não houver viés, esperamos que os códons sinônimos sejam encontrados em **igual quantidade/proporção**
  - Queremos saber **o quão provável é a diferença entre valores observados e esperados** nesse cenário.

- Em outras palavras...
  - Queremos testar o quão provável é uma variável categórica seguir uma determinada distribuição teórica.
  - **Qui-quadrado de aderência**

Resultados dos testes:

O que os resultados significam mesmo?

- Hipótese nula:
  - Não há diferença entre os valores observados e esperados das contagens de codons.
- Dado que a hipótese nula é verdadeira, a probabilidade de obter um resultado onde a **diferença entre valores observados e esperados** é **igual ou maior** aos que eu encontrei é muito baixa (até demais).
- Logo, seria razoável rejeitar a hipótese nula e aceitar uma hipótese alternativa
  - No caso, a hipótese de que há um viés no uso de códons.

Será que os resultados fazem sentido?

Será que os resultados fazem sentido?

Pressupostos qui-quadrado:

- Amostras aleatórias e **independentes**
- Categorias mutuamente excludentes
- Números utilizados formam uma tabela de contingência contendo todos os indivíduos
- n suficientemente grande (pelo menos 20 indivíduos ao todo, valores esperados de pelo menios 5 em cada casela)

É possível que o valor de p baixo esteja apenas associado ao N enorme, mas:

- Não poderia ser um erro na escolha de método estatístico?
- Toda espécie consegue contribuir com um ponto por grupo - Critério de pareamento
  - Meus grupos não são independentes - Isso não quebraria um dos pressupostos?
  - Mas o teste de McNemar serve quando tem uma única variável independente com multiplas categorias?
- Deveria ter tentado comparar as médias das proporções por meio de algum outro teste?
- Teste exato de Fisher: Interessante para quando não tem uma amostragem tão grande
- ANOVA de medidas repetidas...?

Outras perguntas

Pergunta Biológica

A preferência está associada ao número de bases AT presente no códon?

Pergunta experimental

Os valores observados na contagem de códons mitocondriais com 0, 1, 2 ou 3 Adeninas/Timinas são diferentes dos valores esperados caso não haja preferência por nenhuma dessas categorias?

Abordagem:

- Unidade experimental: Espécie
  - Os codons pertencentes à mesma categoria dentro de uma espécie serão usados como réplicas
- Qui-quadrado de aderência
  - 4 categorias: Onde está a diferença?
  - Comparar o quarto grupo (3 A/T) com todos os outros com correção do valor de p por Bonferroni para 3 comparações.

Dois motivos para as réplicas:

1. As unidades experimentais devem ser independentes dentro de cada grupo
2. Os grupos apresentam diferentes numeros de códons: 8, 24, 24, 8. Isso geraria grupos com mais pontos do que outros, o que seria bem bizarro. Em especial para um teste como o qui-quadrado de aderência, que soma valores.

Pitch - Biologia Computacional e Reprodutibilidade

Pitch - Biologia Computacional e Reprodutibilidade

- "We used a custom python/perl script to..."
- "An In-house script was used to..."
- "The 199 mitogenomes were downloaded from NCBI. The complete coding sequence was extracted and codon occurrences were counted for each species."
- Daí você procura o código e ele não está em local algum...
- Você pode contactar o autor do paper e pedir pelo script/jupyter notebook/rmarkdown...
  - Processo demorado: Mais rápido vc mesmo escrever seu programa
  - E isso é péssimo em termos de reprodutibilidade...
- E mesmo que o código esteja disponível, isso ainda não garante que o trabalho seja replicável...

Pitch - Biologia Computacional e Reprodutibilidade

E se focássemos nesses artefatos de pesquisa?

1. Selecionar uma amostra de papers que manipulem dados usando linguagens de programação
- Palavras-chave: **"custom script/program"**, **"jupyter notebook/.ipynb"**, **"Rmarkdown/.Rmd"**, etc...
2. Se por acaso o código não estiver disponível em lugar nenhum, contactar os autores e requisitar o código...
3. Tentar reproduzir as análises computacionais.

Não é uma idéia totalmente original...