Última Aula ME 320 - 29/08/2006

Testes Não-paramétricos

- Os métodos estudados até o momento são válidos na maioria das ocasiões, mas algumas vezes métodos alternativos são necessários.
- Note que para amostras pequenas é necessário assumir que a distribuição populacional não é muito diferente de uma Normal. Em geral, isso não é um problema, mas em alguns casos isso pode ser.

Exemplos em que os testes t não são apropriados são aqueles nos quais:

1. a natureza dos dados implica inevitávelmente numa distribuição extremamente assimétrica ou;
2. os dados não estão numa escala numérica e não faz sentido calcular uma média.

Por exemplo, se tivessemos um escore de dor variando de 1 a 20, a mediana teria uma interpretação mais coerente do que a média.

- Estes métodos não-paramétricos não fazem suposições acerca da distribuição de onde vieram os dados. Eles se baseiam na ordem (postos, ranks) dos dados.
- Embora este procedimento possa parecer melhor, estes métodos são muito menos poderosos, e invariavelmente não fornecem intervalos de confiança.
- Então um conselho é utilizá-los quando as suposições dos outros métodos realmente não parecerem razoáveis.

Amostras independentes

Um biólogo deseja comparar o nÃ°mero médio de besouros capturados numa amostra de 8 armadilhas montadas numa certa floresta, com o obtido numa amostra de 7 armadilhas colocadas numa outra floresta.

As contagens individuais estão listadas abaixo (em ordem numérica):

Amostra 1	8	12	15	21	25	44	44	60
Amostra 2	2	4	5	9	12	17	19	

- Contagens pequenas frequentemente têm distribuições assimétricas, principalmente porque elas devem ser maiores do que zero. Por esta razão, é aconselhável usar um teste não-paramétrico neste caso.
- Para comparar dois grupos independentes (ou não pareados) como estes utiliza-se o teste U de Mann-Whitney.
- Note que as medianas são bem diferentes, mas existe uma certa superposição dos dados, então não é óbvio se existe uma diferença real entre os dois grupos, ou se isto poderia ter ocorrido meramente por acaso.

- O teste de Mann-Whitney primeiro ordena os dados, ou seja, assinala números de 1 a 15 por ordem de tamanho a cada observação, tratando todos os dados como uma grande e única amostra.
- Ele então soma os postos de cada grupo e os compara (com auxílio de uma tabela).
- Quanto maior a diferença nas somas, maior evidência de que existe uma diferença nos tamanhos das observações nos dois grupos.

Usando a tabela adequada para o teste U de Mann-Whitney vemos que neste caso o p-valor é de 0,024. Este p-valor é pequeno então podemos concluir que existe uma diferença estatisticamente significativa nos dois grupos ao nível de 5%.

Portanto, parece existir uma diferença nos números de besouros dependendo do tipo de floresta, e parece existir mais besouros no primeiro tipo de floresta.

Amostras pareadas

Em centros de tratamento de esgoto, amostras podem ser coletadas de duas formas: uma única amostra diária de 2 lts ou amostras pequenas retiradas em 24-horas.

A primeira refere-se a coleta de uma única amostra de 2 lts no mesmo horário diariamente e e segunda baseia-se num esquema de amostragem de 24 horas que retira 1 litro a cada hora.

Um experimento foi conduzido num período de 6 dias registrando-se o número de cistos de Giardia por litro do material.

É de interesse saber se os dados fornecem evidência de que os dois modos de amostragem diferem.

Dia	1	2	3	4	5	6
amostras únicas 2L	100	95	120	175	635	510
amostras 24-horas	145	60	215	670	350	130

Agora podemos usar o teste t pareado, mas como a amostra é muito pequena e os números em cada grupo parecem muito assimétricos, indicando que as diferenças não estarão próximas de uma Normal, um teste não-paramétrico pode ser mais apropriado.

- O teste mais apropriado neste caso é o chamado teste Wilcoxon para dados pareados.
- A forma como ele é feito consiste em primeiro calcular as diferenças das duas medidas em cada par, e então essencialmente testar a hipótese nula de que a diferença mediana é zero.
- As diferenças em valor absoluto (ou em módulo) são ordenadas, ou seja, são assinalados postos Ã s diferenças de 1 a 6. Os postos das observações com diferenças positivas são somados, e os postos das diferenças negativas são somadas.
- Quanto maior for a diferença entre estas somas, maior a evidÃªncia de que existe uma diferença entre os métodos de amostragem.

O p-valor do teste para os nosso dados é 0,917 (obtido de tabela adequada), uma probabilidade muito grande. Isto significa que os dados são consistentes com a hipótese de que não existe diferença nos métodos de amostragem.

Contudo, devemos notar que com tão poucas observações não é de se esperar que existam fortes evidências de uma diferença.

Teste de Kolmogorov-Smirnoff

Sejam $X_1, X_2, \ldots, X_n$ i.i.d. com distribuição F. Queremos testar $H_0 : F = F_0$ versus $H_1 : F \neq F_0$. A estatística do teste será:

$$\sqrt{n} \sum_x |F_n(x) - F_0(x)|,$$

onde

$$\begin{aligned} F_n(x) &= \frac{1}{n}\{\text{número de } X_i\text{'s menores ou iguais a } x\} \\ &= \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(-\infty, x]}(X_i) \end{aligned}$$

é a distribuição acumulada empírica.

Sob H_0 a distribuição da estatística do teste é conhecida, tabelada e não depende de F_0.

Exemplo: Gostaríamos de testar se a hora de nascimentos de bebês é uniforme ao longo do dia. Para isto, 37 nascimentos consecutivos de parto normal na Maternidade foram observados e os dados são: 19:02, 23:08, 3:56, 8:12, 8:40, 12:25, 13:24, 8:25, 14:02, 23:46, 10:07, 13:53, 18:45, 9:06, 15:57, 7:40, 3:02, 10:45, 15:06, 6:26, 16:44, 12:26, 14:27, 23:45, 5:08, 5:49, 6:32, 12:40, 13:30, 00:55, 15:22, 16:09, 19:46, 2:28, 10:06, 11:19 e 16:31.

Neste caso,

$$\sqrt{37}\sup_x |F_n(x) - F(x)| = \sqrt{37}\left|\frac{31}{37} - \frac{1004}{1440}\right| = .85.$$

O valor crítico para $\alpha = 0.10$ é maior que 1.22. Portanto, não temos evidência para rejeitar a hipótese de que os nascimentos ocorrerem uniformemente ao longo do dia.